VOWEL PRODUCTION IN INFANT-DIRECTED SPEECH: AN ASSESSMENT OF HYPERARTICULATION AND DISTRIBUTIONAL LEARNING

George Starling

A thesis submitted in fulfilment of requirements for the degree of

Doctor of Philosophy

to Linguistics and English Language
School of Philosophy, Psychology & Language Sciences
University of Edinburgh

2019

Declaration I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where stated otherwise by reference or acknowledgment, the work presented is entirely my own.

G Starling

Contents

1 Introduction **1**

1.1 Why consider the input and the mechanisms associated with learning? . 2
1.2 Why explore the hyperarticulation hypothesis further? 3
1.3 Why explore distributional learning further? 4
1.4 Defining the learning task 5
1.5 Main contributions and conclusions 7

2 Literature review **9**

2.1 Perceptual attunement 9
2.1.1 Discrimination in infancy 10
2.1.2 Theories of perceptual development 14
2.1.3 Perceptual attunement in English 17
2.1.4 Closing statements on perceptual attunement 19
2.2 Distributional learning 20
2.2.1 Experimental support for this mechanism 21
2.2.2 Modelling studies and distributional learning 28
2.2.3 Closing statements on distributional learning 38
2.3 Hyperarticulation in infant-directed speech 38
2.3.1 Bernstein Ratner: an early view of IDS 39
2.3.2 Vowel space expansion 41
2.3.3 Other measures of vowel hyperarticulation 44
2.3.4 Consonantal distinctions in IDS 46
2.3.5 Multidimensional acoustic data 47
2.3.6 Other accounts of IDS vowel production 49
2.3.7 Closing comments on the hyperarticulation hypothesis 51
2.4 Chapter summary 51

3 Acoustic analysis of F_1 and F_2 in IDS & ADS **53**

3.1 Methodology 53
3.1.1 Materials 54
3.1.2 Data extraction and acoustic analyses 54
3.1.3 Measures of discriminability 59
3.2 Results 63
3.2.1 Area of the vowel space 63
3.2.2 Central tendencies & peripherality 64
3.2.3 Global differences 65
3.2.4 Dispersion 67
3.3 Within-category variance 69
3.4 Degree of overlap 70
3.4.1 Relative orientation of categories 72

3.5 Discussion 75

4 Multidimensional acoustic analysis of IDS & ADS **79**

4.1 Methodology 80

4.1.1 Data extraction and acoustic analyses 80

4.1.2 Measures of discriminability 83

4.2 Results I: F_3 84

4.2.1 Central tendencies 84

4.2.2 Global differences 85

4.2.3 Dispersion 86

4.2.4 Within-category variance 86

4.2.5 Degree of overlap 88

4.3 Results II: patterns of spectral change 88

4.3.1 Central tendencies 88

4.3.2 Dispersion 89

4.3.3 Within-category variance 90

4.3.4 Degree of overlap 90

4.4 Results III: Vowel duration 95

4.4.1 Central tendencies 95

4.4.2 Global differences 95

4.4.3 Dispersion 96

4.4.4 Within-category variance 96

4.4.5 Degree of overlap 96

4.5 Results IV: Multidimensional data 98

4.5.1 Dispersion 98

4.5.2 Degree of overlap 99

4.6 Discussion 99

5 Computational models of perceptual attunement **107**

5.1 Introduction 107

5.2 Methodology 109

5.2.1 Materials 109

5.2.2 Clustering through Expectation Maximisation 109

5.2.3 Multinomial logistic regression 112

5.3 EM clustering 113

5.3.1 Results 113

5.3.2 Interim discussion 121

5.4 Multinomial logistic regressions 123

5.4.1 Results 123

5.4.2 Interim discussion 130

5.5 General discussion 132

6 General discussion **135**

6.1 Thesis summary 135

6.2 The hyperarticulation hypothesis 140

6.3 Distributional learning in infancy 150

6.4 Conclusions and future directions 166

A Confusion matrices **183**

B Regression coefficients **193**

Abstract and lay summary This thesis considers two related research questions that are relevant to vowel production in infant-directed speech (IDS). The first of these research questions determines the extent to which caregivers intentionally clarify vowel distinctions which correspond to differences in word meaning when they address infants in order to promote language learning. The second research question determines the extent to which infants can learn these distinctions by observing statistical regularities in the acoustic input that they receive. In order to observe the modifications that speakers make when they address infants and the effect that these modifications have on infants' learning of the distinctions between vowel sounds, I carried out an acoustic analysis of vowel production data sampled from a large naturalistic corpus of American English caregivers speaking to both infants and adults. This comparative analysis of infant- and adult-directed speech (IDS and ADS) extends previous research in that it applies multiple measures of discriminability to high dimensional acoustic data which details the properties of all of the categories in the system. Though speakers produced a greater dispersion between vowels when they addressed infant learners than adults, their vowel production in IDS was also more variable than ADS. Because of this, the modifications that characterise IDS are not consistent with pedagogical accounts of this register. In order to assess whether infants can identify distinctions from the statistical properties of the input, I applied a series of learning models to the acoustic data that was sampled from each register. Models applied to both IDS and ADS failed to recover the identity of vowels in American English. These results suggest that vowel distinctions cannot solely be learnt by observing the statistical regularities of the input. This research indicates that phonetic data must be examined closely in order to determine the intentions behind the modifications that caregivers make in IDS. The learning models presented here inform current theories of perceptual development by indicating that perceptual learning additionally requires the consideration of infants' early knowledge of multiple levels of linguistic structure. Further to this, this thesis demonstrates that the successful cases of vowel learning in laboratory contexts cannot be trivially generalised to allow learners to discover distinctions in the input that they are exposed to.

Acknowledgements

To my parents Catherine and Mike and my brother Sam, for providing endless support in pursuing research, stimulating an interest in both language and (some aspects of) computer science, and inspiring me to always be confident in my own abilities. The road to the end of my thesis has not been easy and I attribute a lot of the courage that I needed to get here to you three.

To Georgina, for being a wonderful, caring, and understanding partner for all these years and enduring many, many endless train journeys between Edinburgh and London. You never fail to bring a smile to my face, even when things are more than a little stressful and hectic.

To Heather Goad, Andrew Nevins, and Marilyn Vihman, for giving me a strong appreciation of phonological acquisition and the current issues that exist within it. Each of you had vastly different reasons for leading me to pursue developmental questions but the combination of these has resulted in a keen desire to contribute to research in this field.

To Mits, Barbora, and all at ELfLanD, for constantly providing me with a welcoming and warm venue to express and share my thoughts, opinions, suggestions, and rough drafts. Weekly meetings were always a great place to talk through the many small aspects of work that eventually became this thesis. You have both been really strong believers in me since I have known you and in being extremely kind and patient, I don't think I could want for better support throughout this research project.

To James, Pavel, Alice, Bob, Josef, and the rest of the P-Workshop, for continually being a source of engaging discussions and new findings within the field. Each of you have provided me with encouragement to broad the knowledge and skills that I had at the beginning of this project in order to deliver a strong and well-placed contribution to acoustic phonetics as well as developmental phenomena. Also, some of you were involved with me leading tutorials within the department and revealing that teaching was something that I strongly enjoy.

To everyone that I have shared an office with, for a constant but ever-changing peer group to share in this experience with. Your names are far too many to list but thanks to each of you.

1 Introduction

During the first year of an infant's of life, linguistic experience leads to a reorganisation of the learner's perceptual behaviours for speech sounds. Experimental studies have demonstrated that newborn infants show a broad sensitivity to a large number of potentially relevant phonetic distinctions. Discrimination tasks have further provided evidence that infant perception exhibits a shift towards the exclusive discrimination of native language contrasts as they accumulate linguistic experience. Theories of perceptual development have proposed that infants identify native language distinctions by observing statistical regularities in the acoustic input (Maye, Werker, and Gerken, 2002) and have assumed that speech addressed to infants (infant-directed speech or IDS) provides learners with more reliable statistical information about these categories than adult-directed speech (ADS) does (Bernstein Ratner, 1984; Kuhl, Andruski, Chistovich, Kozhevenikova, Ryskina, Stolyarova, Sundberg, and Lacerda, 1997). This thesis contributes to current research by presenting novel empirical data concerning the discriminability of vowels in each of these registers which addresses two related research questions in this domain. The first research question considers the extent to which the properties of IDS are consistent with the enhancement of native language vowel distinctions while the second considers the extent to which a statistical mechanism known as distributional learning can explain perceptual attunement in infancy.

This thesis will present an acoustic analysis of vowel production in American English IDS and ADS as well as a series of learning models which identify the category structure that is apparent in these samples of acoustic data. This thesis will present a renewed analysis of a large corpus of caregivers' spontaneous productions of both IDS and ADS (Bernstein Ratner, 1984). This acoustic data will first enable a discussion of a functionalist explanation of the acoustic properties that have been commonly reported in IDS across the world's languages. Specifically, it has been proposed that speakers modify their production of vowels in this register in order to facilitate the identification of native language distinctions (Bernstein Ratner, 1984; Kuhl et al., 1997). Following Cristia (2013), I will refer to this proposal as the hyperarticulation hypothesis. The evidence that IDS vowels are more discriminable than ADS vowels is contentious, however, since many studies in this domain have applied overly simplistic measures of discriminability to a limited set of the vowels in a given inventory. Acoustic data sampled from IDS and ADS also enables a discussion of the extent to which distributional learning can perceptual attunement in infancy. Experimental tasks have demonstrated that exposure to different types of statistical regularities can alter infants' perception of speech sounds (Maye, Werker, and Gerken, 2002). Though these experiments provide potential evidence for an explanatory mechanism, it remains unclear whether the use of this mechanism allows infants to discover multiple native language distinctions from the inherently variable input that they receive. Computational models which attempt to replicate the use of this mechanism have failed to recover relevant categories from the input without making undue assumptions about the learning task (Antetomaso, Feldman, Miyazawa, Elsner, Hitczenko, and Mazuka, 2016; Feldman, Griffiths, Goldwater, and Morgan, 2013; Frank, Feldman, and Goldwater, 2014; Mooney, 2015).

By considering the hyperarticulation hypothesis and the viability of distributional learning, this thesis explores a one account of the properties of IDS vowel production and one mechanism which may explain the identification of native language distinctions in infancy. Both of these questions aim to address larger research enterprises which respectively aim to describe the properties of speech addressed to children and the mechanisms that explain perceptual attunement in infancy. This thesis should not be viewed as an attempt to exhaustively account for the properties of vowel production in IDS. Similarly, it should not be viewed as an exhaustive account of the emergence of language-specific perceptual behaviours in infancy.

1.1 Why consider the input and the mechanisms associated with learning?

Newborn infants discriminate a broad range of phonetic distinctions regardless of their contrastive status. The emergence of language-specific perceptual behaviours within the first year of life provide evidence of an effect of linguistic experience. This process has been described as perceptual narrowing as development advances through the maintenance of native language contrasts and the loss of non-native distinctions (Tsuji and Cristia, 2013; Werker and Tees, 1984). For vowel distinctions, infant perception becomes language-specific between the ages of six to eight months whereas this change typically occurs at ten months for consonantal distinctions. This thesis will use the x;y format to refer to the age of learners. In this format, the first number indicates the infants' age in years while the second indicates the age in months: for example, 1;8 indicates an age of one year and eight months. Given that perceptual narrowing occurs early in development, theories of perceptual development have favoured mechanisms which are independent on infants' knowledge of other levels of linguistic structure. For example, perceptual attunement is unlikely to depend on lexical information as infants have a limited and sparse receptive vocabulary at this point in development (Caselli, Bates, Casadio, Fenson, Fenson, Sanderl, and Weir, 1995). This precludes the possibility that perceptual attunement can primarily be explained through the identification of minimal pairs: segmental distinctions can be identified as phonological relevant if they result in a change in meaning. The fact that vowel inventories are language specific precludes the possibility that infants have *a priori* knowledge of the number of categories that occur in their native language or of the identity of those categories. Recent theories of perceptual attunement have therefore viewed the emergence of language-specific perceptual behaviours as the result of a domain-general statistical learning process over the acoustic input.

Understanding how linguistic experience shapes perception requires a close consideration of the interaction of the properties of the input and the mechanisms that are available to the infant learner. Investigations of either of these factors depend, at least partially, on the findings associated with the other. For example, analyses which aim to demonstrate that the properties of IDS facilitate learning require a clear understanding of the mechanisms behind this learning process. Distributional learning provides a series of expectations about how the input affects learning as it proposes that perceptual attunement depends on the presence of robust statistical cues to category identity in the acoustic input. Similarly, the viability of distributional learning in infancy depends on the distributional properties of the acoustic input that learners are exposed to. Though the use of this mechanism in laboratory contexts indicate that this mechanism can alter infant perception, it must be demonstrated that these experiments are ecologically valid. Acoustic analyses of IDS must demonstrate that the input that infants are exposed to has similar properties to the experimental stimuli that have been shown to alter infant perception, if this mechanism is to be considered outside of laboratory conditions.

1.2 Why explore the hyperarticulation hypothesis further?

The claim that IDS provides learners with a more discriminable set of acoustic distinctions than ADS forms part of a broader functionalist explanation of the observation that speech addressed to children bears the same properties across a large number of the world's languages (see Cristia, 2013; Saint-Georges et al., 2013; and Soderstrom, 2007 for recent reviews). Comparisons of the the area of the vowel space across IDS and ADS form the primary evidence for the claim that the exaggerated properties of vowel production in IDS may facilitate perceptual learning in infancy (Bernstein Ratner, 1984; Kuhl et al., 1997). This measure of discriminability is defined as the area in acoustic space between the point vowels /i/, /ɑ/ and /u/. The observation of an expanded vowel space in IDS is said to indicate that vowels are more easily discriminated in this register than ADS.

This claim requires further investigation since comparative analyses of these two registers have not always provides strong evidence that vowel categories in IDS are more easily discriminated than those in ADS. Some replications of Kuhl et al. (1997) have failed to observe cases of vowel space expansion in IDS relative to ADS (Bohn, 2013; Burnham et al., 2015; Dodane and Al-Tamimi, 2007; Englund and Behne, 2006). Further to this, vowel space expansion has been criticised for not providing a sufficient description of the distributional properties of IDS and the extent to which they differ from the properties of ADS. Rather than solely indicating the central tendency of a subset of vowels, measures of discriminability must indicate the extent to which the full set of categories overlap in acoustic space. Further to this, these studies have commonly operationalised vowel quality through measures of first two formants. Studies of vowel production and perception in adulthood have indicated that the third formant, vowel duration, and patterns of spectral change also contribute to vowel identity (Hillenbrand, Getty, Clark, and Wheeler, 1995; Hillenbrand, 2013). Because of this, it is possible that acoustic studies in this domain may have either made incorrect conclusions about the intentions behind vowel production in IDS (Eaves Jr. et al., 2016) or they may have overstated the ambiguity of the input that infants receive (Swingley, 2009). The comparative acoustic analysis that will be presented in this thesis will therefore apply multiple measures of discrimnability to all of the phonetic categories of American English in high dimensional data sampled from IDS and ADS. This analysis will also strive to have a maximal degree of ecological validity by considering a corpus of spontaneous speech which greatly resembles the input that infants receive on a daily basis (Bernstein Ratner, 1984).

1.3 Why explore distributional learning further?

The use of distributional learning in infancy has been evidenced through experimental tasks which demonstrated that infants can track the distributional properties of the input and that their perception is altered by what they observed. The first study of this type trained infants who were aged between 0;6 and 0;8 on a stop voicing distinction (Maye, Werker, and Gerken, 2002). In one condition, infants heard tokens of [d] and [t] that were arranged in a bimodal distribution: this condition presented listeners with many tokens were similar to either [d] or [t] and few ambiguous tokens. In the other condition, infants heard a unimodal distribution of tokens which primarily consisted of tokens that were ambiguous between [d] and [t]. Infants continued to discriminate these two consonants in the bimodal condition but failed to discriminate this distinction in the unimodal condition.

The use of this mechanism outside of laboratory contexts depends on an assumption that there is a one-to-one relationship between the number of modes in the frequency distribution of the acoustic signal and the number of distinctions that are found in a system. While some studies of consonantal distinctions show minimal overlap and provide evidence of such a relationship (Lisker and Abramson, 1964; Newman, Clouse, and Burnham, 2001; Sundberg and Lacerda, 1999), American English vowel categories exhibit a considerable degree of overlap in acoustic space (Hillenbrand et al., 1995). Computational models have provided an objective method of determining whether the distributional properties of a given sample of input allow for the identification of the relevant set of phonetic categories. Clustering models which attempt to simultaneously identify the number of categories in a system and the identity of those categories have shown poor performance (Feldman et al., 2013). This model was applied to ADS production data for twelve American English vowels recovered only eight categories which had a poor resemblance to the categories in the input. This suggests that distributional learning does not allow for learners to trivially recover native language distinctions from samples of vowel production data. Computational models have succeeded when the task is simplified either by reducing the number of vowels to be learnt (Vallabha, McClelland, Pons, Werker, and Amano, 2007) or by specifying the number of vowels in the system in advance (de Boer and Kuhl, 2003; Kornai, 1998). The vowel production data from the acoustic analysis of IDS and ADS will serve as input for a series of learning models which will explore two potential reasons for the poor performance of clustering models which greatly resemble the use of distributional learning in infancy. Firstly, these clustering techniques have only ever been applied to data sampled from either IDS or ADS: model performance may be improved by applying these models to IDS, if speakers enhance native language distinctions in this register. Secondly, previous analyses have only considered data where vowel quality was defined by the distribution of F_1 and F_2: model performance may be improved by applying these models to multidimensional data, if additional acoustic dimensions mitigate ambiguity of the input that infants receive (Swingley, 2009). This analysis will further present a series of logistic regression in order to determine the extent to which these additional acoustic dimensions predict individual distinctions in vowel quality across the system.

1.4 Defining the learning task

The discussions above indicate that it is relevant to investigate the properties of the input that infants receive and the mechanisms that may explain perceptual attunement in infancy. It is therefore necessary to provide a clear definition of this learning process. As stated previously, perceptual attunement is a process of reorganisation which is evidenced by a shift from the discrimination of many phonetic distinctions at birth to a set of language-specific perceptual behaviours. The current discussion therefore briefly highlights the empirical evidence for perceptual attunement in infancy and the theories that account for it. Further to this, the current section provides a brief overview of vowel distinctions found in American English as these are the object of the learning task.

Discrimination tasks indicate whether infants are capable of distinguishing a pair of speech sounds at a specific point in time. Experimental tasks which indicate how infants' sensitivity to native and non-native contrasts changes over time therefore provide evidence of perceptual attunement. Specifically, attunement occurs when infants distinguish native distinctions but fail to show a similar sensitivity to non-native distinctions or cases of within-category variation. Broad perception early in development has been evidenced by a study which demonstrated that English infants aged 0;6 discriminated voiceless velar and uvular ejective stops /k', q'/ voiceless dental and retroflex stops /t̪, ʈ/, and voiced labial and alveolar /b, d/ stops (Werker and Tees, 1984). This study provided the initial evidence of perceptual attunement as English infants aged 0;10 were able to discriminate the native English distinction between /b/ and /d/ but showed reduced sensitivity to the non-native ejective and coronal stop distinctions. Discrimination tasks concerning within-category variance also provide evidence of adult-like perceptual behaviour in infancy. When adults and infants aged 0;6 were trained on typical tokens of the category /i/, they perceived atypical tokens of this category as highly similar (Grieser and Kuhl, 1989; Kuhl, 1991). Exposure to atypical tokens of /i/ did not result in the same effect as listeners perceived typical and atypical tokens

	front	central	back
high	i ɪ		u ʊ
mid	eɪ ɛ	ɝ	oʊ ɔ
low	æ	ʌ	ɑ
diphthongs		aʊ aɪ ɔɪ	

Table 1: The vowel categories of American English: in cases where vowels are paired, those to the left of this table are tense while those to the right are lax.

as being distinct in this case. This perceptual effect where exposure to prototypical vowel tokens reduces sensitivity to within-category variation is known as the perceptual magnet effect. The observation of this effect in infancy indicates that learners are sensitive to the internal structure of categories that occur in their native language.

Theories of perceptual development (Kuhl, 1994; Kuhl, Williams, Lacerda, Stevens, and Lindblom, 1992, 2008) describe the properties of the initial perceptual system and propose a set of mechanisms which explain how infants' perceptual sensitivities change throughout development. These theories have primarily stated that infant perception is shaped by phonetic properties of the input and thus attribute the use of distributional learning in infancy a central role (Maye, Werker, and Gerken, 2002). Additionally, these theories describe how perceptual attunement interacts with infants' emergent knowledge of the lexicon, phonotactics and phonology of their native language. The interaction between different levels of linguistic structure is described as a mutually beneficial relationship. For example, lexical information, such as the recognition of minimal pairs in the input, can help infants to establish the existence of native language distinctions. Heightened sensitivity to native language distinctions also facilitates the recognition of different instances of the same lexical item or phonotactic sequences. Though broader theories of development allow for a greater understanding of perceptual change in infancy, it should be highlighted at this point at these factors have limited relevance to the relative discriminability of across registers and the viability of distributional learning infancy.

For the most part, this thesis will take an agnostic view of the distinctions that infants induce through the use of distributional learning and that caregivers enhance in IDS as previous analyses have applied the term *phonetic category* to both individual phones and more abstract phonemic units. To this end, the current analysis will consider considers the phonemic distinctions that are found in the inventory of American English. As indicated in table 1, this system consists of twelve monophthongs and three diphthongs and the contrasts between these vowels are evidenced by the existence of minimal pairs. The current discussion of vowel production and perceptual development does not address the status of individual phones or allophonic variants within this system. The consideration how these units are realised in the acoustic signal and the mechanisms by which infants process them in order to discover phonemic distinctions is beyond the scope of the current discussion.

When considered in the context of typologically common vowel inventories, the inventory of American English is comparatively large. The distinctions between the twelve monophthongs in this system primarily differ in height and backness. Tense-lax distinctions further separate pairs of vowels which are similar in terms of height and backness. Since every front vowel is unrounded and the majority of back vowels are rounded, lip rounding is generally not considered to be contrastive in American English. This feature may, however, further distinguish distinctions in backness. Rhoticity is unique to the central vowel /ɝ/.This system has a set of three raising diphthongs: both /aɪ/ and /ɔɪ/ become more advanced throughout their duration but differ in their initial quality. The vowel /aʊ/ moves towards a high, back quality throughout its duration. The unstressed vowel, [ə], will not be considered in this analysis as it is not phonemic in American English.

1.5 Main contributions and conclusions

The main findings of the thesis presented in the subsequent chapters are the following. In chapter 2, I review current literature in order to define perceptual attunement and motivate the two research questions which I address in the thesis. The first of these considers the extent to which the properties of IDS vowel production facilitate the recognition of vowel categories in infancy while the second considers the extent to which distributional learning accounts for this learning process. Chapter 3 presents an extensive acoustic analysis of both infant- and adult-directed speech, consisting of formant measures for all fifteen American English vowel categories. Though this analysis indicates that vowels have greater dispersion in IDS than ADS, it further highlights that within-category variance is greater in this register than ADS. I demonstrate that variance-sensitive measures provide evidence against the hyperarticulation hypothesis as IDS vowels show a greater degree of overlap than ADS. Chapter 4 extends this comparative acoustic analysis by considering multidimensional acoustic data from each register. Measures of dispersion and variance again indicate that IDS vowel categories do not show a lesser degree of overlap than those in ADS. Multidimensional acoustic data therefore does not provide strong evidence of enhancement in IDS than formant measures do. Chapter 5 presents clustering models and logistic regressions which further confirm this lack of enhancement in IDS. The performance of models applied to IDS data do not generally outperform those applied to ADS data. Clustering models which replicate distributional learning fail to recover an appropriate number of categories. These models therefore predict that perceptual attunement cannot be explained through the observation of the statistical regularities of the acoustic input. The results of the logistic regressions indicate that both IDS and ADS provide learners with ambiguous cues to American English vowel quality distinctions.

These findings have implications for current accounts of the properties of infant-directed speech and of perceptual attunement in infancy. The lack of enhancement that was observed in IDS challenges the claim that the quality of vowels in this register reflect the caregiver's intention to facilitate the identification and processing of relevant distinctions. This empirical investigation highlights that high within-category variance relative to ADS should be considered as a definitive property of IDS vowel production. Variance-sensitive measures must be adopted in this domain in order to capture the relative discriminability of vowels across registers and document how the distributional properties of the input differ across registers. The adoption of such measures prompts a further analysis of the properties of this register and a further consideration of how caregivers' intentions shapes vowel production in IDS. The statistical models in this thesis challenge the current view that distributional learning is a central explanatory mechanism behind perceptual attunement. These models indicate that exposure to the statistical properties of IDS and/or ADS does not lead to adult-like perceptual behaviours. Instead, distributional learning over this input would reduce learners' sensitivity to native language distinctions. By demonstrating that statistical mechanisms do not sufficiently explain perceptual attunement, this thesis prompts a more holistic view of development which incorporates infants' emergent knowledge of the lexicon and phonotactics of their native language.

2 Literature review

This thesis explores two hypotheses that are relevant to the identification of native language vowel distinctions in infancy. The first research question of this thesis concerns the extent to which the phonetic properties of IDS facilitate the identification of these distinctions. If the properties of vowel production in IDS benefit infant learners, this would provide a functionalist explanation for the observation that speech addressed to infants bears similar properties across the world's languages. The second research question assesses the extent to which perceptual attunement can be explained through the use of distributional learning in infancy. This statistical mechanisms provides a potential explanation for the emergence of language-specific perceptual behaviours within the first year of life. This literature review consists of three sections that motivate and contextualise these two related questions. The first section of this chapter provides an overview of early speech perception and establishes the relevant perceptual phenomena that have been observed in infancy. By summarising theories of perceptual development, this section also highlights the relationship between the properties of the input that infants are exposed to and the learning mechanisms that are available to them. The second section in this chapter discusses the experimental evidence that demonstrates how infant perception can be altered by the distributional properties of the input that they are exposed to. These experiments indicate the specific properties of input that may induce perceptual attunement as well as the limits of this mechanism in laboratory contexts. This section also highlights a series of learning models that replicate the use of this statistical mechanism in order to determine whether this mechanism allows for categories to be recovered from samples of acoustic data. The third section of this chapter discusses the phonetic properties of IDS and reviews the evidence that caregivers adjust the realisation of categories in this register in order to highlight native language distinctions. It also focusses on the acoustic measures that have been adopted in this domain in order to describe the relative discriminability of vowels across registers. These measures should ideal demonstrate that IDS provide learners with the relevant statistical information about the vowel inventory of their native language.

2.1 Perceptual attunement

Speech perception in adulthood is influenced by the set of distinctions that are present in the phonology of a speaker's native language. Adult speakers show categorical responses for native language distinctions rather than showing gradient behaviour. When native speakers of English are presented with a continuum between /ɹa/ and /la/, they show a sudden shift from identifying the tokens as instances of /ɹ/ to identifying them as instances of /l/ (Miyawaki et al., 1975). These behaviours are language-specific since speakers show poorer discrimination when presented with non-native distinctions. Speakers of Japanese, a language where /ɹ/ and /l/ are not contrastive, did not distinguish any of the tokens from the same continuum. Adult speakers also show language-specific perceptual behaviours when they are presented with instances of within-category variance. Adult speakers of English were trained on tokens of /i/ and then tested on their ability to detect variants of this vowel (Grieser and Kuhl, 1989; Kuhl, 1991). Listeners were either trained on prototypical exemplars of this vowel or poor instances of this category. In the prototypical condition, adults showed a reduced sensitivity to within-category variation. Adults in the non-prototypical condition perceived the same variants more easily. Cases where native speakers are biased towards the prototypical members of a category are referred to as the perceptual magnet effect. Atypical tokens are 'drawn in' to the centre of the category, reducing adults' sensitivity to these tokens.

Infant discrimination tasks have revealed that these language-specific behaviours emerge within the first year of an infant's life (see Tsuji and Cristia, 2013, for a review). At birth, infants show fine-grained sensitivity to a broad range of potentially relevant acoustic distinctions. Native language experience leads to a reorganisation of this initial perceptual system: native language distinctions become sharpened while non-native distinctions are less easily discriminated. Language-specific patterns of discrimination emerges at around 0;6 for vowel distinctions and at around 0;10 for consonantal distinctions. This process of reorganisation has been referred to as perceptual attunement. The current discussion will consider the discriminatory capabilities that infants have at birth and changes in perceptual sensitivity during the first year of life. This discussion will also highlight theories of perceptual development that explain how infant perception is shaped by the properties of their native language. In addition to describing the initial system and the process of attunement to the native system, these theories must account for developmental patterns other than the maintenance of native distinctions and the loss of all others.

2.1.1 Discrimination in infancy

Evidence of categorical perception has been observed in English infants between the ages of 0;1 and 0;4 (Eimas et al., 1971). At both ages, subjects discriminated voiceless /p/ and voiced /b/ labial stops from another but did not discriminate between instances of the same category. Infants of a similar age also discriminated the vowel distinctions /i, u/ and /i, ɑ (Trehub, 1973). Categorical perception has also been observed for vowel distinctions. Though English infants aged 0;2 discriminated /i/ from /ɪ/, they failed to detect within-category variantion since they could not discriminate each vowel from intermediate tokens (Swoboda et al., 1978).

The first evidence of perceptual attunement was observed in a series discrimination tasks with English infant subjects aged 0;6, 0;8, and 0;10 (Werker and Tees, 1984). These tasks provided evidence that the broad sensitivity sense in early development is followed by the loss of non-native distinctions. Infants aged 0;6 and 0;8 discriminated a native distinction between voiced bilabial /b/ and alveolar /d/ stops, a Nɬeʔkepmxcin distinction between voiceless velar /k'/ and uvular /q'/ ejective stops and a Hindi distinction between voiceless dental /t̪/ and retroflex /ʈ/ stops. Infants' perceptual behaviour at 0;10 indicated attunement with their native language as these older infants were only sensitive to the English native distinction.

Tasks involving vowel distinctions have demonstrated that the same pattern occurs earlier in development. Infants from Canadian English backgrounds were capable of discriminating the German distinctions between /u:/ and /y:/ and between /ʊ/ and /ʏ/ at the age of 0;4 but failed to discriminate either of these contrasts at 0;6 (Polka and Werker, 1994). Evidence of the perceptual magnet effect provides further evidence that infant perception becomes attuned to their native language. When English infants aged 0;6 were trained on prototypical exemplars of /i/, they showed reduced sensitivity to within-category variance (Grieser and Kuhl, 1989; Kuhl, 1991).

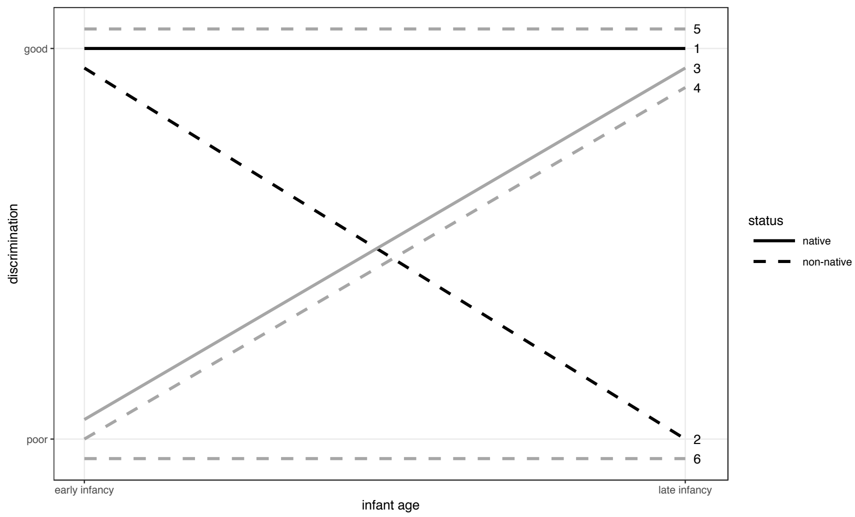

Figure 1: Attested patterns in perceptual attunement. The expected patterns (black) are the maintenance of native distinctions (1) and the loss of non-native distinctions (2). Mazuka, Hasegawa, and Tsuji (2014) report four further patterns (grey): as development progresses, both native and non-native distinctions may become more easily discriminated (3, 4) while non-native distinctions remain either easily (5) or poorly perceived (6).

The following section will address the additional patterns of change that are highlighted in Mazuka, Hasegawa, and Tsuji (2014) and illustrated in figure 1. These additional patterns represent phenomena that theories of perceptual development must account for. As well as revealing the complexity of perceptual attunement, these patterns provide further insights into the properties of the early perceptual system and the influence of native language input. Cases where discrimination improves throughout development have been observed for both native and non-native distinctions. These patterns indicate that neonates are not equally sensitive to all possible distinctions and that development cannot solely be viewed as the loss of non-native distinctions. Developmental theories must allow for linguistic experience to have a positive effect on discrimination and predict the set of native and non-native distinctions that show improvement. Further to this, these theories must address the observation that some non-native distinctions are not lost, even in the absence of relevant experience.

Improvement for native distinctions Discrimination in early infancy is poor when the two relevant sounds have a high level of acoustic or articulatory similarity. For example, infants from Canadian English and Filipino language backgrounds did not discriminate a low salience distinction between alveolar /na/ and velar /ŋa/ nasals at 0;4 or at 0;8 (Narayan, Werker, and Beddor, 2010). Since /n/ and /ŋ/ were presented in onset position, this was only a native distinction for Filipino infants. English does have a contrast between these sounds but only in codas. Discrimination tasks with English and Filipino infants aged between 0;10 and 1;0 showed a positive effect of native language experience since only Filipino infants were sensitive to this distinction. This effect of salience was supported by the observation that infants from both backgrounds and age groups consistently discriminated a more salient distinction between labial [m] and alveolar [n] nasals. Similar behaviour has been observed for a distinction between the voiced alveolar stop [d] and the voiced dental fricative [ð] (Polka, Colantonio, and Sundara, 2001). Neither French nor English infants showed sensitivity to this distinction between 0;6 and 1;0. Since successful discrimination was only observed in a control group of English adults, this study did not indicate a positive effect of native language experience within the first year of life. Both of these experiments indicated cases where infants showed a lack of sensitivity to low salience non-native distinctions throughout development.

It follows from these results that theories of perception must account for perceptual salience and explain how greater exposure to the linguistic input can improve a learner's sensitivity to native language distinctions. Positive effects of linguistic experience are not limited to low salience distinctions. Both English and Japanese infants were sensitive to a distinction between /la/ and /ɹa/ between 0;6 and 0;8 (Kuhl et al., 2006). A further set of discrimination tasks with infants aged 0;10 to 1;0 indicated that English infants showed greater sensitivity while Japanese infants showed reduced sensitivity to this distinction. Improved discriminatory capabilities may depend on infants' emergent knowledge of the lexical and phonotactic properties of their native language. The delayed discrimination of /d/ and /ð/ has been attributed to the observation that /ð/ is phonotactically restricted in English and that it most frequently occurs in function words (Polka, Colantonio, and Sundara, 2001).

Improvement without exposure In contrast to the cases where non-native distinctions remained imperceptible through development, infant discrimination tasks have also indicated that infants can show an increase in their sensitivity to non-native distinctions. Japanese infants were tested on three German contrasts (/i, e/, /u, o/, and /u, y/: Mazuka, Hasegawa, and Tsuji, 2014). Infants' behavioural responses at 0;4 again indicated that salience can affect early discrimination as learners were only sensitive to the distinction between /u/ and /y/. At 0;10, infants showed increased sensitivity to the distinction between /i/ and /e/. Since these German vowels did not occur in the Japanese input, this discrimination task presented a case of improvement without exposure. As this older set of subjects did not discriminate /u, y/ or /u, o/, these distinctions respectively demonstrate an expected pattern of decline and a lack of sensitivity throughout development. Infants' improved discrimination of non-native /i/ and /e/ was attributed to the fact that exposure to Japanese /i/ and /ɛ/ provided learners with experiences that was relevant to the German distinction between /i/ and /e/. Improvement without exposure was not observed for German /u/ and /o/ as these vowel were not sufficiently similar to Japanese /ɯ/ and /o/. Both of the German vowels have stronger rounding than the Japanese back vowels. Improvement without exposure has also been observed where American English infants showed an increase in sensitivity to Finnish /i/ and /y/ between the age of 0;7 and 0;11 (Cardillo, 2010). Though /y/ does not occur in the English vowel inventory, fronted variants of native /u/ may have provided English infants with relevant experience.

Maintenance of non-native distinctions Discrimination tasks have indicated that certain non-native distinctions remain perceptible throughout development. For example, English infants consistently discriminate a distinction between voiceless labial and alveolar ejectives, [p'] and [t'], between 0;6 and 1;0 (Best, 1994). Infants remain sensitive to these ejective stops due to their similarity to the native stops /p/ and /t/. Cases of maintenance have also been observed for salient distinctions between sounds that are dissimilar from those that occur in the native inventory. Infant learners of American English continually discriminated dental [|] and lateral clicks [||] between the age of 6 and 14 months (Best, Roberts, and Sithole, 1988). The lack of decline in discriminability suggested that infants' native language experience did not influence their perception of these sounds. Alternatively, infants may have processed clicks as non-speech sounds.

Asymmetries in early perception Investigations of the early perceptual system have revealed asymmetries in the early perceptual system as well as cases where there was an absolute lack of sensitivity. These asymmetries occur when infants only discriminate the distinction when they were trained on one specific vowel and tested on the other. English infants aged 0;4 and 0;6–0;8 showed asymmetric discrimination for two German contrasts (/u, y/ and /ʊ, ʏ/: Polka and Werker, 1994). Infant subjects only discriminated these distinctions when they were trained on the back vowels, /u/ and /ʊ/. When trained on front vowels, infants successfully discriminated these non-native distinctions. The existence of this asymmetry was only transitory since English infants aged between 0;10 and 1;0 did not discriminate either distinction regardless of the vowel that they were trained on.

These asymmetries have been observed for the German distinction between /u/ and /y/ and the English distinction between /ɛ/ and /æ/ (Polka and Bohn, 1996). These patterns were observed across four sets of subjects: this study considered English and German infants aged 0;6–0;8 as well as an older cohort of infants aged 0;10–1;0 from both language backgrounds. These discrimination tasks did not indicate an effect of age or language background on perception. Infants that were trained on the peripheral vowels /u/ and /æ/ failed to discriminate the distinction while those that were trained on /y/ and /ɛ/ showed successful discrimination. One interpretation of these results is that the experimental paradigm failed to capture a change in infants' sensitivities. If it was successful, however, this paradigm may have indicated that not all contrasts undergo a decline in sensitivity throughout development. These results may also have provided insights into how native language experience shapes the perception of non-native contrasts. English listeners may have been biased towards perceiving German /u/ since it was comparable to the English phoneme /u/ while German /y/ was comparable to fronted allophones of English /u/. German infants may have shown a lack of an effect of linguistic experience as neither /æ/ or /ɛ/ were sufficiently similar to native German vowels.

2.1.2 Theories of perceptual development

Discrimination tasks have indicated that the developmental patterns that theories of perceptual attunement must account for. These theories must explain positive effects of linguistic experience as well as accounting for declines in sensitivity to non-native distinctions. Changes in infants' discriminatory capabilities must be linked to infants' experience with the acoustic properties of the input or their emergent knowledge of other levels of linguistic structure. Such positive effects of experience must also explain why infants show an improved ability to distinguish certain non-native distinctions without predicting that all non-native distinctions follow this developmental pattern. These theories must also describe the state of the initial perceptual system, accounting for the effects of salience and the biases that have been observed in early perception. As well as describing the infants' discriminatory capabilities, these theories of development also consider the emergence of abstract phonemic categories. The current discussion will primarily consider the native language magnet theory (NLM: Kuhl, 1994; Kuhl et al., 1992, 2008) as an explanatory model of perceptual development and further discuss the Perceptual Assimilation Model (Best, 1994), PRIMIR or Processing Rich Information from Multidimensional Interactive Representations (Werker and Curtin, 2005), and the Natural Referent Vowel framework (Polka and Bohn, 2011). These four approaches should be viewed as separate implementations of a single approach rather than as competing theories. Each theory provides a unique focus on different aspects of perceptual attunement. These approaches have been described as an implicit two-stage process (Dillon, Dunbar, and Idsardi, 2013). In the first stage, infants establish individual phones by drawing statistical inferences over the acoustic input. In the second stage, learners establish abstract phonemic units by grouping this initial set of phonetic categories together on the basis of their emergent knowledge of other levels of linguistic structure. This stands in contrast to a single-stage model where infants identify phonemes and the relations between the allophones contained within them by simultaneously attending to the acoustic signal and the higher level units of linguistic structure.

Native language magnet theory Native language magnet theory (NLM: Kuhl, 1994; Kuhl et al., 1992, 2008) provides an account of how infants' experience with their native language in the first year guides the emergence of language-specific perceptual behaviour. The broad discrimination of both native and non-native distinctions in the early perceptual system is attributed to infants' general auditory processing capabilities. Though these perceptual capabilities allow for the discrimination of many potentially relevant distinctions, infant perception can be crude and some low salience distinctions may not be perceived robustly. For example, infants' performance in discrimination tasks is poorer than that of adult listeners. This framework highlights the statistical properties of the input as the key factor that drives perceptual attunement in infancy. As native language experience accumulates, infants begin to identify native language categories by observing the statistical regularities in the input. Regions of acoustic space with a high frequency of tokens correspond to category centres while regions of low density indicate category boundaries. Experimental evidence of distributional learning (Maye, Werker, and Gerken, 2002) and the facilitative properties of infant-directed speech (Kuhl et al., 1997) are viewed as empirical support for this statistical approach. Language-specific perceptual behaviours emerge since the statistical properties of the input 'warp' infants' perception of acoustic space. The perceptual space between native categories expands while regions near category centres contract. Since this modulates the similarity of tokens, categorical perception can be explained as a heightened sensitivity to perceptually distal acoustic units. Conversely, the contraction of acoustic space near category centres causes the namesake magnet effect as cases of within-category variance become indistinguishable from prototypical tokens.

Though this model focuses on how the statistics of the acoustic input shape auditory perception in infancy, it further describes how perceptual development interacts with other levels of linguistic structure. Language-specific perceptual behaviour facilitate the recognition of lexical items and phonotactic patterns in the input. This relationship is mutually beneficial since the lexicon and phonotactics may influence the perceptual system. The recognition of minimal pairs, for example, can sharpen infants' discrimination of a specific distinction. Under this framework, the properties of the input are allowed to affect the perceptual system throughout development. However, continued exposure to homogenous input will cause the system to stabilise as this type of experience will only reinforce distinctions that have already been established.

Perceptual Assimilation Model This framework discusses how infants' perception of non-native distinctions can indicate the emergence of native language categories (Best, 1994). Changes in infants' sensitivity to non-native distinctions are therefore explained through infants' exposure to their native language. Cases of decline and improvement without exposure are explained by considering the mappings that infants establish between native and non-native distinctions. As native categories emerge, non-native speech sounds become assimilated to them on the basis of similarity. This approach assumes a close link between speech production and perception. This approach assumes that infants attend to gestural constellations, rather than acoustic properties, in order to establish native categories. Regardless of whether an acoustic or articulatory approach is taken, this framework provides important insights into perceptual development by describing how native language categories affect the perception of non-native distinctions.

Under this model, cases of decline occur when both non-native sounds are assimilated to a single native category. English infants show a lack of sensitivity to the Hindi distinction between dental /t̪/ and retroflex /ʈ/ stops since both sounds are assimilated to the emerging native alveolar /t/ (Werker and Tees, 1984). The maintenance or improvement of non-native distinctions indicate cases where each non-native category is assimilated to a different native category. This framework also predicts that non-native distinctions remain discriminable when both sounds are sufficiently dissimilar from any native sounds that occur in the input. These dissimilar speech sounds cannot be assimilated to any native categories and thus no perceptual reorganisation occurs. Infants aged between 0;6 and 1;0 show robust discrimination of dental [|] and lateral clicks [||] and continue to do so into adulthood because these consonants cannot be assimilated to any native category. Thus, the retention of non-native distinctions should be viewed as providing relevant insights into perceptual development rather than evidence that the infant perceptual system undergo reorganisation.

Natural Referent Vowel framework The natural referent vowel framework describes the properties of the initial perceptual system. Specifically, it proposes that there is a bias that favours the perception of the extreme acoustic and articulatory properties of the three point vowels, [i], [ɑ], and [u]. The point vowels serve as primitive categories which provide salient points of reference and indicate the limits of the vowel space. Infants can identify less peripheral vowels by comparing them to these referents. This preference for peripheral vowels accounts for an asymmetry that has been observed in early perception. When infants are trained on peripheral vowels, they show poor discrimination of internal vowels: training on internal vowels, however, does not affect infants' discrimination of peripheral vowels. A preference for peripheral vowels can be explained in a parallel fashion to the preference for native categories that is observed with the perceptual magnet effect. Since this bias occurs in the early perceptual system, this framework views cases where peripheral vowels are discriminated symmetrically as evidence of perceptual attunement.

PRIMIR PRIMIR provides a broader account of speech perception that addresses perceptual attunement alongside later stages of acquisition such as the emergence of receptive vocabulary and phonemic distinctions (Werker and Curtin, 2005). These developmental patterns are described through a series of interactions between planes that represent the different sources of information that learners extract from the linguistic input. This dynamic approach allows infants' perceptual capabilities to vary dependent on the task that they are facing. While infants reliably distinguish two sounds in a discrimination task, they may not demonstrate this sensitivity in word learning (Stager and Werker, 1997). The General Perceptual plane allows infants to identify phonetic distinctions in the input. Individual exemplars are clustered on this plane and native language categories are established through the use of distributional learning. The Word Form plane represents infants' emergent knowledge of the lexicon and consists of sequences of phonetic categories. This knowledge allows infants to identify the types of variation that are permissible within specific lexical items. These word forms must later be linked to specific concepts. Infants' emergent knowledge of word forms allows for a further Phonemic plane to develop. Abstract categories emerge as infants isolate the cases of variability that correspond to category distinctions from indexical or coarticulatory variants. PRIMIR views the task demands, any perceptual biases and the infants' development level as filters on the information that learners can access from these planes. This filter explains the variable performance that can be observed throughout development and across tasks. This permits the existence of a mutually beneficial relationship between word learning and perceptual attunement while still allowing for performance to differ across discrimination tasks and tasks involving word learning.

2.1.3 Perceptual attunement in English

The current discussion has considered a series of discrimination tasks that highlight the phenomena that must be explained by theories of perceptual development. These theories highlight how the initial perceptual system changes through exposure to the statistical properties of the input. A primary goal of this thesis is to determine the extent to which distributional learning can explain the shift in infants' sensitivity. Since this question will be investigated by considering the vowels of American English, it is necessary to review discrimination tasks indicate how English infants process native vowel distinctions. As illustrated in figure 1, it is not trivial to assume that all native distinctions are well perceived at birth and maintained throughout development. The results of discrimination tasks inform the types of mechanism that are involved in this learning task. Low salience distinctions indicate that infants require further exposure to the acoustics of the input or that they must depend on other levels of linguistic structure. Alternatively, cases where discrimination is robust or improves in early development suggest that these vowels are well distinguished in acoustic input.

Early investigations into English native vowel distinctions demonstrated that infants distinguished /i/ and /ɑ/ at 0;1–0;4 and that this distinction was maintained at 0;6 (Kuhl, 1979; Trehub, 1973). Robust discrimination is expected since these two vowels are maximally distant from each other in acoustic space. Many studies have therefore used this distinction as a control to validate experimental methods (Polka and Bohn, 1996) or to explore whether infant perception is robust to differences in pitch or speaker identity (Kuhl, 1979, 1983; Marean, Werner, and Kuhl, 1992). Infants aged 0;6 that were trained on the distinction between /i/ and /ɑ/ discriminated these sounds when they were spoken by different speakers or had a different pitch contour. Poorer performance was seen with a less salient distinction between /ɑ/ and /ɔ/. Half of the infant subjects trained on this distinction did not generalise it across speakers (Kuhl, 1983). Early literature demonstrated evidence that English infants aged 0;2 perceived the distinction between /i/ and /ɪ/ categorically. These infant learners were not sensitive to within-category variance. This insensitivity to within-category variants of /i/ was also observed in infants aged 0;6 (Grieser and Kuhl, 1989; Kuhl, 1991).

More recent experiments with older infants have provided limited support for the maintenance of native language distinctions. Canadian English infants aged 0;8 discriminated a distinction between /ɪ/ and /eɪ/ (Pons, Sabourin, Cady, and Werker, 2006b) in a distributional learning paradigm (Maye, Werker, and Gerken, 2002). This paradigm features two training conditions which are intended to facilitate or attenuate infants' discrimination of the given contrast. Since infants discriminated these vowels in both conditions, this paradigm presented evidence of robust discrimination. Monolingual and Spanish-English bilingual infants failed to discriminate /eɪ/ and /ɛ/ at the age of 0;4 (Sundara and Scutellaro, 2011). Since infants aged 0;8 from both backgrounds discriminated this distinction, this task indicated that this distinction may not be salient and requires further exposure to the linguistic input. As discussed previously, English infants showed asymmetric discrimination of /æ/ and /ɛ/ throughout the first year of life (Polka and Bohn, 1996). Given that German infants showed a similar pattern of discrimination, these results did not present transparent evidence of perceptual attunement.

Given that many English distinctions have not been tested consistently across multiple age groups, it is hard to make broad generalisations about the status of distinctions throughout the first year of life. It should be noted that patterns of maintenance have been observed across many native distinctions from other languages (Cheour et al., 1998 on Finnish /i, y/ from birth to 0;3; Jansson-Verkasalo et al., 2010 on Finnish /ø, e/ from 0;6 to 1;0; Sebastián-Gallés and Bosch, 2009 on Catalan /e, ɛ/ and /o, u/ from 0;4 to 0;8; Benders, 2013 on Dutch /ɑ, aː/ from 0;11 to 1;3). Evidence of infants' sensitivity to a larger sample of English vowel distinctions has been observed in studies that consider amount of phonetic detail in learners' early lexical representations (Bergelson and Swingley, 2017; Mani and Plunkett, 2010). As highlighted under PRIMIR (Werker and Curtin, 2005), word learning tasks may require a different set of skills than discrimination tasks and thus these results should not be compared directly. The amount of detail in infants' receptive vocabulary was tested by presenting learners with objects that were either labelled correctly or labelled with a vowel mispronunciation. If infants are sensitive to native vowel distinctions and exploit this knowledge in word learning, the mispronounced forms should not be accepted as valid labels. At the age of 1;0, British and American English learners looked longer at objects when they were correctly labelled in comparison to mispronounciations. This indicated that infants were sensitive to the set of distinctions indicated in tables 2 and 3. These do not provide transparent evidence of contrast maintenance as the status of these distinctions before 1;0 remains unreported. Additionally, this lexical task does not independently demonstrate infants' discriminatory capabilities since acoustic and lexical factors were not isolated in this task.

Distinction	Word	Actual	Mispronunication
ɔ, u	ball	bɔːl	buːl
ɑː, ɔː	bath	bɑːθ	bɔːθ
æ, ɑ ː	cat	kæt	kɑːt
ʌ, ɛ	cup	kʌp	kɛp
ɒ, ɑː	dog	dɒg	dɑːg
ʊ, ɔː	foot	fʊt	fɔːt
æ, ʌ	hand	hænd	hʌnd
ɪ, ɛ	milk	mɪlk	mɛlk

Table 2: Word forms and their corresponding mispronunciations in Mani and Plunkett (2010), indicating the distinctions that British English infants detected at 1;0.

2.1.4 Closing statements on perceptual attunement

Discrimination tasks have indicated that infants' perceptual sensitivities become aligned with their native language within the first year of life. Theories of perceptual attunement state that infants identify relevant categories through the use of statistical inferences over the phonetic input. Though development can broadly be described as the maintenance of native distinctions and the loss of all others, the observation of other patterns of development have prompted a need for more nuanced theories of development. Perceptual salience may affect the discriminability of distinctions in the initial system. For low salience distinctions, the emergence of robust discrimination may require greater exposure to the acoustic input or generalisations from other levels of linguistic structure. Infants' emergent knowledge of these levels of linguistic structure allow for the establishment of abstract phonemic categories which can be used in later phonological development. Because of this, these theories have been labelled as presenting an implicit two-stage approach (Dillon, Dunbar, and Idsardi, 2013) where

Distinction	Word	Actual	Mispronunciation
æ, oʊ	apple	æpəl	oʊpəl
æ, u	banana	bənænə	bənunə
ɑ, ɪ	bottle	bɑɾəl	bɪɾəl
ʊ, ɑ	cookie	kʊki	kɑki
ɪ, ɔ	ear	ɪɹ	ɔɹ
aɪ, eɪ	eyes	aɪz	eɪz
eɪ, aʊ	face	feɪs	faʊs
ʊ/i, eɪ	foot, feet	fʊt, fit	foʊt
ɛ, ɑ	hair	hɛɹ	hɑɹ
æ, ʌ	hand	hænd	hʌnd
u, aʊ	juice	dʒus	dʒaʊs
ɛ, u	leg	lɛg	lug
ɪ, ʌ	milk	mɪlk	mʌlk
aʊ, ɪ	mouth	maʊθ	mɪθ
oʊ, æ	nose	noʊz	næz
oʊ, eɪ	yoghurt	joʊgəɹt	jeɪgəɹt

Table 3: Word forms and their corresponding mispronunciations in Bergelson and Swingley (2017), indicating the distinctions that American English infants detected at 1;0.

infants first identify a set of phones and then later group these phones into phonemes on the basis of their emergent knowledge of other levels of linguistic structure.

These theories of perceptual attunement attribute a central role to the distributional properties of the input and the inferences that learners can draw from these regularities. Assessments of the viability of distributional learning require a detailed analysis of the phonetic input and computational models that replicate the outputs of this learning process. Though only a limited set of English distinctions have been tested, discrimination tasks have generally provided evidence that is consistent with the robust discrimination of native language distinctions in early infancy and the maintenance of these distinctions throughout the first year. Studies which have demonstrated that infants are capable of discriminating a broad range of native distinctions cannot attribute these perceptual effects to distributional learning: infants' performance on these tasks necessarily depended on their receptive vocabulary.

2.2 Distributional learning

A primary goal of this thesis is to determine the extent to which perceptual attunement can be explained through the statistical regularities present in the input. Such an assessment requires a review of the empirical evidence that supports the use of this mechanism in infancy. Infants have been shown to be capable of tracking the statistical properties of the input and altering their perceptual behaviours to align with what they observe (Maye, Werker, and Gerken, 2002). These results reflect other studies that have shown that passive exposure to statistical patterns enables infants to identify word boundaries (Saffran, Aslin, and Newport, 1996), phonotactic patterns (Chambers, Onishi, and Fisher, 2003) and constraints on word order (Gomez and Gerken, 1999). The use of this mechanism has important implications for development theories since it provides an explicit link between the properties of the input and infants' perceptual behaviour. Thus, it must be demonstrated that this mechanism can reliably influence infants' perception of a broad range of potentially relevant distinctions. When assessing these experiments, it is necessary to consider the specific statistical properties that affect infant discrimination. For learners to use this mechanism outside of laboratory contexts, the input infants receive must contain these statistical regularities. Furthermore, if distributional learning plays a major role in perceptual attunement, it follows that hyperarticulation in IDS should ensure that learners are exposed to statistical regularities that correspond to native language categories.

Though theories of perceptual development consider distributional learning alongside other mechanisms, this mechanism is unique in that it allows infant learners to discover relevant phonetic categories independent of their knowledge of the adult lexicon. Approaches that depend on the existence of minimal pairs are unlikely to succeed since infants' receptive vocabulary is sparse during the first year of life. English infants understand an average of 36 at 0;8 and this increases to 86 at 1;0 (Caselli et al., 1995). The early lexicon contains very few minimal pairs which are required to motivate the existence of phonemic distinctions: the words *hot* and *hat* form the only minimal pair in the first fifty items that infants recognise. Further to this, experimental tasks have indicated that infants aged 0;8 cannot learn minimal pairs as labels for novel objects (Stager and Werker, 1997). The observation that this ability emerges at 1;2 suggests that lexical approaches to perceptual attunement are unsuitable.

2.2.1 Experimental support for this mechanism

Maye, Werker, and Gerken (2002) was the first study that demonstrated that the infant speech perception is affected by exposure to statistical regularities in the acoustic signal. English infants aged 0;6 and 0;8 were trained on a distinction between prevoiced [da] and voiceless unaspirated [ta]. This distinction is not contrastive in English which instead contrasts plain [d] and aspirated [$\mathrm{t^h}$] stops. The training stimuli were sampled from an eight-step continuum of tokens between these two sounds. These tokens differed both in the amount of voicing prior to the stop (negative voice onset time or VOT) and the trajectory of the first two formants after the stop's release.

Infant subjects were allocated to one of two familiarisation conditions. Each condition presented subjects with different types of statistical information by differing the frequency with which each stimulus was presented. The unimodal condition presented more tokens from the centre of the continuum while the bimodal condition present more typical tokens of [da] and [ta]. Figure 2 indicates the specific frequencies of each stimulus in each condition. The unimodal condition was analogous to languages where voicing is not contrastive while the bimodal condition was analogous to relevant native input. Both familiarisation conditions presented infants with six blocks consisting of 24 tokens paired with a unrelated visual stimulus. In each block, infants heard sixteen training items and eight fillers which consisted of tokens of [ma] and [la].

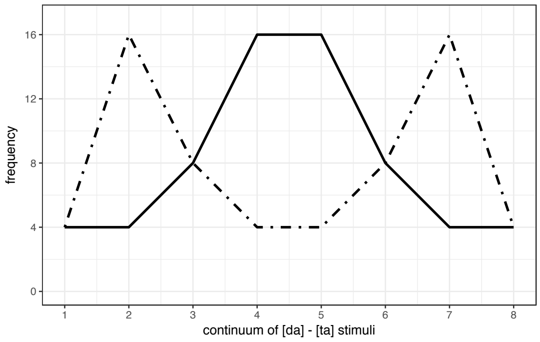

Figure 2: The frequency of each token in the continuum in the bimodal (dashed line) and unimodal (solid line) conditions in Maye, Werker, and Gerken (2002). This figure highlights how the number of modes in the input shapes infant perception.

Trial Type	Stimuli
Non-Alternating	3333...
	6666...
Alternating	1818...
	8181...

Table 4: Test trial types in Maye, Werker, and Gerken (2002)

After familiarisation, infants were presented with eight test trials in which they listened to auditory stimuli while looking at a checkerboard pattern. There were two types of test trial as indicated in table 4. Non-alternating trials repeatedly presented a single stimulus while alternating trials alternated between the two extremes of the continuum. At the start of each trial, a visual stimulus briefly attracted the infant's attention and their mean looking time to the checkerboard was the dependent variable. Differences in looking times across the two trial types indicated their discriminatory capabilities. Since the familiarisation phase contained a variety of tokens, infants were predicted to show a novelty effect of longer looking times for non-alternating trials.

In the bimodal condition, infants looked longer at the stimulus in non-alternating trials in comparison to alternating trials. Infants' looking times did not differ across trial types in the unimodal condition. These results had two possible interpretations: either the bimodal condition facilitated discrimination or the unimodal condition attenuated discrimination. The latter result is consistent with the observation that perceptual attunement features the loss of non-native distinctions. Further to this, discrimination tasks indicate that English infants discriminated this voicing distinction between the ages of 0;6 and 0;8 (Pegg and Werker, 1997).

Distributional learning of consonantal distinctions Subsequent replications of Maye, Werker, and Gerken (2002) have attempted to demonstrate that this mechanism available for multiple distinctions in infant subjects from a range of age groups and language backgrounds. Studies which fail to replicate this result indicate the limits of this mechanism, suggesting that some populations or distinctions are not affected the statistical regularities in the input. The discussion will further assess the methodological differences between these experimental studies in order to identify the specific statistical properties that alter perception. Further to this, this review will consider the ecological validity of this learning mechanism and its implications for learning outside of laboratory contexts.

Distributional learning has been demonstrated in English infants aged 0;10 using the same contrast between prevoiced [da] and unaspirated [ta] (Yoshida et al., 2010). With this older cohort, differences in perceptual behaviour were only observed across trial types when infants were exposed to a longer bimodal familiarisation condition. Though infants required twice the number of tokens to demonstrate this perceptual effect, the training that subjects received was brief in comparison to months of native language experience that infants receive. This study also demonstrated that a decline in perception occurred when infants are exposed to a flat distribution where all tokens had an equal frequency. The attenuation of perceptual sensitivity in these experiments therefore occurred in the absence of clear distributional information rather than as a specific effect of exposure to a unimodal distribution.

Other replications of distributional learning have extended this perceptual effect to other distinctions and acoustic cues. One such replication considered the distinction between prevoiced /d̥a/ and voiceless unaspirated retroflex /ʈa/ stops that are contrastive in Hindi (Capel et al., 2011). Dutch infants aged from 0;9 to 0;11 discriminated this contrast when they received bimodal input and failed to do so in the unimodal condition. This mechanism also affects non-native fricative place contrasts. English infants aged between 0;4 and 0;6 showed differential discrimination of the contrast between the Polish voiceless retroflex /ʂ/ and alveolopalatal /ɕ/ fricatives across unimodal and bimodal conditions (Cristia et al., 2011). Distributional information about the place distinction was cued by the spectral centre of gravity (CoG) of the fricative as well as the formant transitions into the following vowel.

A modified variant of this paradigm has indicated that exposure to distributional information for one distinction affected infants' perception of a similar distinction (Maye, Weiss, and Aslin, 2008). This study considered the distinction between the prevoiced and voiceless unaspirated alveolar stops, [d] and [t], and a parallel distinction between the velar stops, [g] and [k]. English infants aged 0;8 were allocated to one of four familiarisation conditions. In two conditions, the infants exclusively heard alveolar tokens during familiarisation while the other two conditions presented learners with velar tokens. As in the original paradigm, infants were presented with either a bimodal or a unimodal condition for one of the two places of articulation. At test, infants heard alternating and non-alternating trials featuring the voicing distinction that they did not hear in familiarisation. A comparison of looking times across trial types indicated that infants in the bimodal condition discriminated the novel voicing distinction while infants in the unimodal condition did not. Infants therefore generalised the presence or absence of a voicing distinction across places of articulation.

Distributional learning of vowel distinctions Given that this thesis primarily focusses on vowel distinctions, it is important to demonstrate that exposure to statistical regularities also affects infants' perception of vowel categories. Initial studies demonstrated that this paradigm did not affect infants' discrimination of distinctions in vowel length or quality (Pons, Mugitani, Amano, and Werker, 2006; Pons, Sabourin, Cady, and Werker, 2006). For the length contrast, Canadian English infants aged 0;6 were trained on tokens of [ɛ] and [ɛː] that only differed in vowel duration (Pons et al., 2006a). For the quality distinction, Canadian English infants aged 0;8 were trained on /e/ and /ɪ/ (Pons et al., 2006b). In both of these experiments, infants showed a significant difference in looking times across trial types in both familiarisation conditions. The observation of robust discrimination indicated that exposure to a bimodal distribution did not facilitate discrimination nor did exposure to a unimodal distribution attenuate infants' sensitivity.

The use of ERP (event-related potential) methods has provided evidence that distributional learning was available for vowel distinctions in Dutch infants aged between 0;2 and 0;3 (Wanrooij, Boersma, and Benders, 2015). The use of neural rather than behaviour responses allowed for this effect to be tested in a younger cohort of infants. Exposure to statistical cues affected how these infants perceived an English distinction between /ɛ/ and /æ/. This distinction depended on the position of the first two formants: the vowel /ɛ/ had lower frequency for F_1 and a higher frequency for F_2 in comparison to /æ/. The familiarisation phase lasted approximately twelve minutes and presented learners with either a bimodal or unimodal distribution consisting of 900 unique tokens. As in the original study, the bimodal condition presented a greater number of typical tokens while the unimodal condition presented a greater number of tokens that were phonetically intermediate between /ɛ/ and /æ/. The testing phase used an oddball paradigm which repeatedly presented subjects with one vowel as a standard. Infants' neural responses to the presence of the other vowel (or the deviant) indicated whether they discriminated this distinction. This testing phase lasted approximately 30 minutes since infants did not have to remain awake to provide neural responses. Infants showed greater mismatch negativity when presented with the deviant in the bimodal condition in comparison to the unimodal condition. Thus, infants had stronger discriminatory capabilities after exposure to bimodal input. A post-hoc test further indicated that infants in the bimodal condition showed stronger discriminatory capabilities when /ɛ/ was the standard. This finding contradicts the predictions of the PAM (Polka and Bohn, 2011) since relevant language experience should eliminate the existence of this type of perceptual asymmetry.

Study	L1	L2	Distinction
Gulian, Escudero, and Boersma (2007)	Bulgarian	Dutch	/ɑ, aː/; /i, ɪ/
Escudero, Benders, and Wanrooij (2011)	Spanish	Dutch	/ɑ, aː/
Wanrooij (2015)	Spanish	Dutch	/ɑ, aː/
	Dutch	English	/ɛ, æ/

Table 5: Details of a set of studies which have demonstrated that distributional learning can be used by adult learners to identify non-native vowel contrasts.

Though evidence of distributional learning has only been observed in this isolated case for infant learners, exposure to distributional information has been demonstrated to affect adults' perception of non-native vowel distinctions as indicated in table 5.

Limitations and ecological validity These experiments provide important empirical support for developmental theories that view the statistical properties of the acoustic input as a driving factor behind perceptual attunement in infancy. Though the investigation of a broader set of distinctions would be of merit, replications of Maye, Werker, and Gerken (2002) have indicated that this mechanism affects both vocalic and consonantal distinctions indexed by different acoustic dimensions. Experiments have also indicated that this mechanism is available for the majority of the first year of life. Observations made about one distinction can also be generalised to other analogous distinctions. Though certain studies have failed to replicate the effects of distributional learning (Pons et al., 2006a,b), these results do not necessarily indicate that the use of this mechanism is limited. Robust discrimination across conditions is consistent with the idea that these distinctions were salient to infant learners. Alternatively, input in which only a single acoustic dimension contained statistical regularities may not have been sufficient to alter infants' perception (i.e. vowel duration for [ɛ], [ɛː]; F_2 for /e/, /ɪ/).

Given that distributional learning has been shown to be a general and reliable mechanism, considerations of the ecological validity of these results may provide a more valid critique of this learning mechanism. Specifically, it is necessary to consider the extent to which the experimental stimuli resembled native language distinctions in infants' linguistic input. During familiarisation, infant learners learnt about a single phonetic distinction from a continuum of eight unique tokens that were arranged to have specific distributional properties. These stimuli were presented in monosyllabic or disyllabic word forms which formed a minimal pair and were produced in isolation. By contrast, natural language presents learners with a larger number of phonetic categories and so multiple phonetic distinctions must be processed simultaneously. Each token in the input will be acoustically unique and instances of a single speech sound may occur in multiple lexical, prosodic and phonotactic contexts. The experimental stimuli can therefore be viewed as a condensed version of linguistic experience that eliminates many sources of acoustic variability. Infants would have to isolate a specific distinction and control for any sources of variance in order to accumulate experience with a similarly homogenous set of tokens.

Even if it is assumed that infants are exposed to this type of input, the ecological validity of these experiments further depends on the assumption that the linguistic input presents learners with the same statistical regularities as those of the experimental stimuli. Assessing the similarity of the information across these contexts requires the consideration of two factors. Firstly, it is necessary to identify specific aspects of the stimuli that induced behavioural differences across the unimodal and bimodal conditions. Once these relevant properties have been established, it is necessary to determine whether these same properties occur in the typical input that infants receive.

These experiments have primarily attributed this effect to the number of modes in each frequency distribution. Familiarisation tasks where infants were trained on a larger set of unique acoustic tokens indicated that infants depend on region of high frequency density rather than tracking the raw frequency of a specific token (Wanrooij, 2015). Despite this, this paradigm has been open to alternative interpretations since the number of modes was not the only property that differed across conditions. The standard deviation of the relevant acoustic dimension across tokens in the bimodal condition was greater than the standard deviation across tokens in the unimodal condition, enabling an explanation where high variance input facilitated discrimination. In unimodal and bimodal conditions with identical standard deviations, Spanish adult subjects successfully discriminated a Dutch contrast between /ɑ/ and /aː/ in both conditions (Wanrooij, Boersma, and Benders, 2015). Adult subjects failed to discriminate these vowels when they were heard classical music in the familiarisation phase. Paradigms with infant learners have not provided support for this alternative explanation (Yoshida et al., 2010). English infants showed differences in perceptual behaviour after exposure to bimodal and flat conditions, even though the standard deviation of tokens was equal across these two familiarisation conditions (Yoshida et al., 2010).

The stimuli used in these paradigms presented infants with one-to-one relationship between modes in the frequency distribution and contrastive sounds. The existence of individual modes in a frequency distribution is defined by three factors: the central tendency of each category, their variance and their frequency. Certain samples of consonantal distinctions in laboratory speech have indicated the existence of a bimodal distribution for pairs of categories (Lisker and Abramson, 1964; Newman, Clouse, and Burnham, 2001; Sundberg and Lacerda, 1999). As in the bimodal condition, categories in these samples of speech were distal from one another and had minimal within-category variance. By contrast, phonetic studies of the vowel system of American English have indicated that there is a considerable degree of overlap between these categories (Hillenbrand et al., 1995). Cases of overlap, such as those indicated in figure 3, hinder the identification of individual modes which is required for distributional learning to succeed.

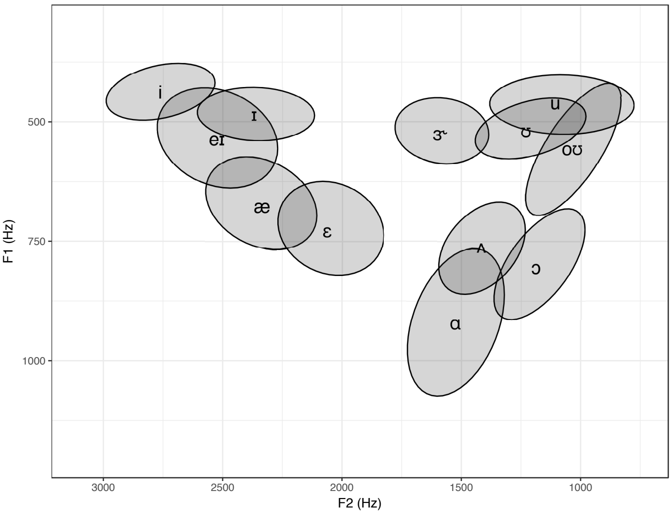

Figure 3: The distribution of the mean of F_1 and F_2 for twelve American English vowels across 48 female speakers in Hillenbrand et al. (1995). Each symbol indicates category identity and the grand mean across speakers while the ellipses are 80% confidence regions. This figure highlights cases of overlap which hinder the use of distributional learning.

These familiarisation conditions also presented learners an equal number of tokens from each category. Low frequency categories in the input are problematic since these categories may not have an individual mode in the frequency distributional. An analysis of phonological length in Japanese IDS found that duration was a reliable predictor of this distinction, indicating that the short and long vowels are well dispersed relative to their variance (Bion et al., 2013). Despite this, vowel duration was unimodally distributed in this sample of input. Since 94% of Japanese vowels were phonologically short, the infrequent long vowels did not manifest as an individual mode.

This discussion of distributional learning in infancy has indicated how the observation of statistical regularities in the input can alter the perception of speech sounds. Though this mechanism has been applied to a range of contrasts and appears robust, the viability of this mechanism crucially depends on the quality of the information that is present in the input. Specifically, this mechanism depends on the existence of a one-to-one relationship between the number of distinctive speech sounds that are found in a language and the number of modes in a frequency distribution of relevant acoustic dimensions. Therefore, considerations of this mechanism must observe the relative position, variance, and frequency of categories in the input. Despite the fact that theories of perceptual development attribute a central role to this mechanism, it is not trivial to assume that the input that infants receive contains reliable distributional information. Relatedly, the assessments of the hyperarticulation hypothesis which assess differences in the discriminability of categories across registers must also measure the relative position, variance and frequency of category across registers. If caregivers do intentionally clarify native language distinctions, categories in IDS should show a minimal degree of overlap in order to enable the use of this mechanism.

2.2.2 Modelling studies and distributional learning

Computational models that replicate distributional learning in infancy provide an objective method of determining the quality of the distributional information in a sample of acoustic data. These models have implemented the use of distributional learning to the exclusion of all other learning mechanisms. By attempting to identify the central tendency, variance and frequency of native language categories, these computational models make explicit predictions about the type of generalisations that learners may draw from a given sample of acoustic input. Models that successfully recover the category structure support the use of distributional learning in infancy while poor model performance indicates that native language categories cannot be identified though the use of distributional learning alone. Comparing model performance across IDS and ADS can also provide insights into the relative discriminability of vowels across these two registers. Cases where models applied to IDS data outperform those applied to ADS data would indicate that the acoustic properties of IDS facilitate the use of distributional learning in infancy.

The goals of these models must be considered closely when assessing their relevance to theories of perceptual attunement. Broadly speaking, these computational models have attempted to demonstrate whether distributional learning can identify phonetic categories. However, this term has applied ambiguously to refer to both phonemes and individual phones and the distinction between these two units of representations has commonly been collapsed. The majority of models in this domain have been evaluated with regard to their ability to identify phonemes on the basis of the distributional properties of the acoustic input. This characterisation of the learning task has been highlighted as problematic in a discussion of how computational models correspond to theories of perceptual development (Dillon, Dunbar, and Idsardi, 2013). Ultimately, the goal of this task is to provide learners with a set of phonologically relevant units. English infants must establish that /t/ is distinct from /d/. Additionally, they must establish the conditioning environments in which /t/ is realised as the allophones [t$^{\text{h}}$], [ɾ], and [t̚]. Models which view phonemes as the output of distributional learning would leave learners insensitive to these allophonic distinctions. Thus, these models do not closely align with the single-stage approach where learners first establish a set of phones which are later grouped together to form abstract phonemic categories. These models are also said to be an inappropriate method of replicating the single-stage model. By considering distributional learning as the sole mechanism behind this task, these models do not implement the simultaneous observation of lexical and phonotactic information which is required in order to identify allophones as subunits of individual phonemes.

Three separate methodologies can be identified across the modelling studies in this domain. This discussion will first address studies that have assessed the status of distinctions using logistic regressions. This supervised technique indicates the extent to which a series of continuous variables affect the probability of a response to a categorical variable. In this case, these models determine how each acoustic dimension predicts category identity. Since these models are trained on labeled data, this approach should not be viewed an attempt to replicate the infant learning task. Instead, the regressions indicate the performance of an optimal observer which fully exploits the statistical regularities of the input. This discussion will also consider two sets of studies that use clustering techniques that resemble distributional learning in infancy. Statistical clustering is a unsupervised process in which data points are assigned to groups on the basis of similarity. Clustering approaches can be further distinguished into two methods dependent on whether the number of categories in system is given or not. One class of clustering models aim to identify the parameters that describe a fixed number of categories. Another class of clustering models incorporates the identification of the number of categories in the system into the statistical learning process. Since infants do not have any *a priori* knowledge of the number of distinction in their native language, the latter approach more closely resembles the use of distributional learning in infancy.

Logistic regressions Logistic regressions indicate the quality of distributional information by determining the extent to which a series of continuous acoustic dimensions affect the likelihood of a response to a single categorical distinction. A regression coefficient, β, indicates the strength of each acoustic dimension as a predictor of identity. These models will have numerically large coefficients for dimensions which show little or not overlap while ambiguous dimensions will have smaller coefficients. A Wald test which considers the estimate and standard error of these coefficients can be used to determine whether a given acoustic dimension significantly affects the probability of a response for the relevant distinction.

Regression analyses have indicated that both American English and Japanese IDS contain reliable distributional cues to a pair of distinctions that are similar across these languages (American English, /i, ɪ/, /ɛ, e/; Japanese /i, iː/, /ɛ, ɛː/: Werker et al., 2007). This analysis operationalised vowel quality as the difference between F_2 and F_1 as well as measuring the duration of each vowel. The regression indicated that formant values were the only significant predictor for the English distinctions while vowel duration was the only predictor for the Japanese distinctions. Duration was also found to be a significant predictor of phonological vowel length in Japanese IDS (Bion et al., 2013). Note, however, that these results were not sensitive to the differences in frequency. These studies may therefore have overstated the reliability of the distributional information that is present in the input.

Multinomial logistic regressions have also been applied to multiple vowel distinctions in American English IDS and ADS (McMurray et al., 2013). In contrast to the binary distinctions presented above, multinomial logistic regressions assess the probability of responses to a categorical factor with more than two levels. This analysis can be

		predicted		
		i	ɑ	u
	i	13	0	7
actual	ɑ	0	20	0
	u	5	0	15

Table 6: An example of a confusion matrix containing fictional data for the classification of three vowels. This matrix indicates that tokens of /ɑ/ were always identified correctly while tokens of /i/ and /u/ had greater confusability.

viewed as a series of separate logistic regressions where multiple categorical outcomes are compared to a single reference category. This regression explored the extent to which first three formants predicted eight different vowel qualities: /iɪ/, /eɪɪ/, /æ/, /ɝ/, /ɑɪ/, /ʌ/, /oʊ/, and /ɑɪ/. This study did not report the coefficients for each of these acoustic dimensions. Instead, model performance was assessed by predicting the identity of categories in an unseen data set. The accuracy of this predictive technique was assessed through the construction of a confusion matrix. As indicated in table 6, a confusion matrix establishes correspondences between the actual identity of categories in the unlabeled data and the identities that were predicted by the model. Classification accuracy is defined as the number of true positives – cases where the model predicted category identity correctly – divided by the total number of categories that the model considered, or the sum of the matrix. Greater performance was observed in ADS (72.9%, 95% CIs = [66.0, 79.2]) in comparison to IDS (69.9%, 95% CIs = [63.0, 76.7]), suggesting that the properties of IDS did not support language acquisition. Measures of accuracy for individual categories indicated that peripheral vowels such as /i/, /æ, and /oʊ/ showed greater accuracy than interior vowels such as /ɝ/, /ʌ/, and /ɑɪ/.

Clustering models with a fixed number of categories Statistical clustering has been used to model distributional learning as a process by which vowel tokens are assigned to a series of categories. Gaussian mixture modelling is a common clustering technique which assumes the vowel system can be modelled as a series of Gaussian or normal distributions. Each vowel token is a member of one of K Gaussian distributions, where K corresponds to the number of categories in the system. Three parameters define each of these Gaussian distributions. The mixing probability, π, indicates frequency of the given category or the number of tokens assigned to it. The mean, μ, locates the category in acoustic space. The variance, Σ, defines the limits of the category. In the cases of multidimensional distributions, these limited are defined by a covariance matrix. Given that each of these parameters affect the quality of the distributional information, this process of statistical clustering strongly resembles distributional learning in infancy. This method can therefore be used to determine how these properties affect the viability of this approach. Gaussian mixture models locate a maximally likely set of parameters through an iterative process known as expectation maximisation. Since each category is modelled as a probability distribution, the model's likelihood is maximised when each data point is assigned to the distribution that it is closest to. This approach ensures that tokens are assigned to clusters on the basis of similarity.

The current discussion will focus on Gaussian mixture models that aim to estimate the parameters of a fixed number of categories within a given data set. In this case, the value of K is specified as the number of categories in the sample of acoustic data. These models will be contrasted with Gaussian mixture models that view the identification of the number of categories in the system as part of the learning task. In short, these models must identify an optimal value for K as well as estimating values for π, μ and Σ.

An early study in this domain (Kornai, 1998) used k-means clustering to successfully identify the categories apparent in the American English vowel production data collected in Peterson and Barney (1952). This model identified mean formant values that were comparable to the values that the original study reported for ten monophthongs (/i/, /ɪ/, /ɛ/, /æ/, /ɝ/, /ɑ/, /ʌ/, /ɔ/, /ʊ/, /u/). An clustering analysis applied to the English point vowels (/i/, /ɑ/, /u/) in IDS and ADS (de Boer and Kuhl, 2003) assessed model performance by considering the model's estimates for the central tendency of each category. IDS models identified the central tendencies that more closely resembled the properties of the input more closely than the ADS models did. This was interpreted as support for the facilitated properties of this register. A second clustering analysis of same IDS and ADS production data operationalised vowel quality using MFCCs (mel frequency cepstral coefficients) rather than formant values (Kirchhoff and Schimmel, 2005). This alternative method was assessed using measures of classification accuracy and indicated poorer model performance in IDS (93.5%) than ADS (95.5%). Despite the fact that this study did not replicate the advantage for IDS, both registers showed near-ceiling performance.

This technique has been presented as support for the claim that prosodically prominent vowels in American English IDS provide learners with robust distributional information. An expectation-maximisation model that classified all tokens of /i/, /ɑ/ and /u/ had a classification accuracy of 91.5% (Adriaans and Swingley, 2017). A parallel model that exclusively considered tokens that occurred in prosodic focus showed an improvement in classification accuracy to 96.7%. Though models applied to /i/, /ɪ/ and /ɛ/ showed poorer performance with an accuracy of 65.6%, performance still improved to 71.4% when focussed tokens were considered in isolation. Table 7 provides the results of clustering models which were applied to four further vowel pairs. Similar effects were observed in an earlier study that considered the IDS productions of a single speaker (Adriaans and Swingley, 2012). The model classified all tokens of /ɛ/, /æ/ and /ɑ/ with accuracy of 64.5% and showed improved performance of 73.4% when only prominent tokens were considered. The differences in model performance observed across these sets of vowels were consistent with the idea that cases of overlap

	all tokens		focused only	
i, ɪ	78.9	[78.3, 79.5]	86.7	[86.5, 86.9]
ɪ, ɛ	70.2	[69.3, 71.2]	72.1	[71.4, 73.0]
ɛ, æ	68.0	[67.4, 68.7]	69.6	[69.1, 70.0]
æ, ɑ	79.6	[79.0, 80.3]	86.3	[86.2, 86.5]

Table 7: Measures of classification accuracy with 95% confidence intervals for the clustering models in Adriaans and Swingley (2017). These measures indicate that performance improved when only focussed tokens were considered. Despite this, performance was poorer than that of models which considered the three point vowels.

limit the viability of distributional learning. Point vowels showed the highest accuracy since these vowels are maximally distal from one another. The poorer performance seen in models applied to groups of phonetically similar vowels suggested native language categories could not be trivially recovered through the use of distributional learning, even when prosodically prominent vowels were considered in isolation.

Clustering models with an unknown number of categories Infants do not have *a priori* knowledge of the number of distinctions found in their native language. In order to capture this aspect of the learning task, modelling studies have adopted clustering techniques which estimate an optimal value for K as well as the parameters of each cluster. Infinite Gaussian mixture models and model selection through the Bayesian information criterion (BIC) are two approaches that do not require the number of clusters to be specified in advance.

Infinite Gaussian mixture models estimate K by starting with an arbitrarily large set of categories and iteratively eliminating these from the model until an appropriate value for K is found. When the model is initialised, each of the tokens in the data is sequentially assigned to a category. The first token in the data is assigned to its own category. Each subsequent token then has a probability of being assigned either to an existing category or a new one. Thus, as each token is added, the value of K increases. This model limits the initial value of K through a bias that states that tokens should be assigned to categories with a large number of members. These initial assignments are considered in order to generate an initial value for K and parameter values for each of the K categories in the model. The model then attempts to improve on these initial estimates through an iterative process. Each iteration first reassigns each token to its most likely category. Since categories are probability distributions, this will ensure that each category consists of a set of maximally similar tokens. These reassignments have the potential to reduce the value of K since categories that have no tokens assigned to them are eliminated from the model. The other step of this iterative process updates the parameters of each category dependent on the new tokens that were assigned to them. The model converges once this iterative process can no longer makes changes to category assignments or parameters that increase the model's likelihood.

A variant of this model has been implemented where individual tokens are added to the model incrementally (McMurray, Aslin, and Toscano, 2009; Vallabha et al., 2007). This stands in contrast to models which process all of the tokens in a single batch. These incremental models are initialised with an arbitrarily large number of categories and the parameters of these categories are updated as each new token is added. Competition between categories ensures that the model converges on an optimal number of categories. Each token is assigned to its most likely category as it is added, ensuring that each category consists of a set of maximally similar tokens. As well as changing the mean and variance of the selected category, the addition of new tokens alters the frequency of each category in the system. The frequency of the most likely category is increased while the frequency of lower likelihood categories is reduced. This alteration of frequency ensures that this model converges on an optimal value for K since it eliminates low frequency categories from the model.

BIC-based model selection provides an alternative method for locating an optimal value for K. Rather than trying to converge on an optimal value of K, this method generates a series of Gaussian mixture models that each have a different of value for K and then selects one optimal model from amongst these. For each value of K in the specified range, expectation maximisation is used to estimate a set of maximally likely category parameters for a given data set. Models are evaluated using the BIC, which is defined as the difference between a model's complexity and its likelihood. Likelihood, $\hat{L}$, is defined by the similarity of each data point to the category that it is assigned to. Complexity is defined by the number of parameters that a model estimates, n. Models with higher values for K have a greater complexity since a value for π, μ and Σ must be estimated for each category. The model with the minimum value for the BIC is selected as having the optimal value for K.

$$\mathrm{BIC} = \ln(n)k - 2\ln(\hat{L})$$

Evaluating models through the BIC is necessary since an appropriate value for K cannot be located solely through comparisons of likelihood or complexity. Model selection on the basis of complexity would always favour the model with the minimal value for K. Conversely, model selection on the basis of likelihood would always select the model with the largest values for K. Likelihood is maximised when data points are assigned to compact categories. Selecting the model that has the smallest value for BIC allows for a trade-off between these two factors. Overly simple models will be penalised for their lower likelihood while increasingly complex models must be justified by an accompanying increase in likelihood. In short, the BIC-optimal model will have good fitness relative to its complexity.

Though the number of categories provides one indicator of performance, computational models have also been assess through pairwise measures of accuracy. Since these models do not necessarily converge on a correct value for K, correspondences cannot be established between the categories in the input those predicted by the model. Confusion matrices are therefore of limited use. Pairwise measures of accuracy assess model performance by comparing whether pair of vowel tokens were members of the same category or not, both in the actual data and in the model's allocations. True positives (TP) are cases where the model allocated a pair of tokens to the same class that genuinely were instances of the same vowel category. False positives (FP) occur when the model considers a pair to be the same that should be distinct. False negatives (FN) are pairs that the model considers to be distinct that should be the same. F-scores are defined as the harmonic mean of two accuracy measures, precision and recall.

$$\text{Precision} = \frac{N_{TP}}{N_{TP} + N_{FP}} \qquad \text{Recall} = \frac{N_{TP}}{N_{TP} + N_{FN}}$$

$$\text{F-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

These computational approaches have closely approximated the use of distributional learning in infancy by applying infinite Gaussian mixture models to entire inventories. These models have been applied to the twelve monophthongal vowels of American English (/i, ɪ, eɪ, ɛ, æ, ɝ, ɑ, ʌ, ɔ, oʊ, ʊ, u/: Feldman et al., 2013). These categories were learnt from values of the first two formants resampled from vowels read in isolation and produced in a single phonological context (Hillenbrand et al., 1995). The frequency of each vowel category was defined by frequencies reported for parental speech in CHILDES (Li and Shirai, 2000; MacWhinney, 2000). The clustering model typically recovered eight categories and showed poor performance (pairwise F-score = .52). Similar models have been formant values sampled from Glaswegian English sociolinguistic data (Mooney, 2015) and Japanese and American English IDS data (Antetomaso et al., 2016). The model applied to Glaswegian vowels again underfit the data, recovering six clusters with a pairwise F-score of .47. These clusters showed a poor correspondence with the nine vowels of this system (/i, ɪ, e, ɛ, a, ʌ, ɔ, o, ʉ/). Cases of overfit have also been reported in this domain (Antetomaso et al., 2016). Models that considered Japanese (F score = .22) and American English IDS data (F score = .13) recovered a larger number of categories than those found in the input. The English model recovered twenty categories rather than the intended twelve while the Japanese model recovered 22 categories rather than five vowel qualities that further differ in length (/i(ː), ɛ(ː), a(ː), o(ː), ɯ(ː)/).

Improved model performance has been seen when models were applied to a subset of vowel categories. Gaussian mixture models were applied to the formant and duration data sampled from four front vowels in Japanese (/i, iː, ɛ, ɛː/) and American English (/i, ɪ, eɪ, ɛ/) IDS production data (Vallabha et al., 2007). This acoustic measures were originally collected in Werker et al. (2007). These models typically located four categories and the median classification accuracy of these runs was 91.1% for Japanese and 92.7% for American English. BIC-based clustering models also successfully recovered two categories /aː/ and /ɑ/ when applied to distributions of F_2 and duration from

Dutch IDS data (Benders, 2013). The same modelling technique has indicated the potential availability of a covert contrast in final stop voicing in Dutch (Kirby, 2014). Solutions with two categories were selected as optimal when models were presented with the durational measures of the stop burst, the preceding vowel, the stop closure and the voicing in the closure.

class	categories	acoustic measures
vowels	/i/, /y/, /u/	F_2, F_3
glides	/j/, /ɥ/, /w/	F_2, F_3
fricatives	/s/, /ʃ/	spectral CoG
stops	/k/, /g/	VOT

Table 8: The set of distinctions sampled from French IDS production data in Moeng (2016)

Models that considered subsets of a language's consonantal categories have also generally recovered an appropriate number of categories. A Gaussian mixture model successfully identified the contrast between English plain and aspirated stops (McMurray, Aslin, and Toscano, 2009). This model considered values of voice-onset time that were sampled from read speech where stops were produced in initial position (Lisker and Abramson, 1964). 97% of these models converged on a two category solution which accurately estimated the central tendency of each category. A series of BIC-based clustering models (Moeng, 2016) were applied to French IDS data consisting of the four sets of categories indicated in table 8. The three high vowels were recovered well as indicated by measures of recall. These models showed poorer performance for the consonantal distinctions. Tokens of /k/ were split across two categories and a fourth spurious category was identified in the glide data. The model applied to the fricative data correctly identified two categories, measures of recall indicated that performance was poor. Though 97% of the tokens of /ʃ/ were assigned to a single category, this category also contained many tokens of /s/. These misclassifications were highlighted by a value of recall for /s/ of 58%.

Clustering models with additional levels of structure Clustering models have shown poor performance when applied to entire vowel inventories. Because of this, these models have suggested that distributional learning alone cannot explain perceptual attunement in infancy. This conclusion only holds true if the learning task is viewed as a process by which infants attempt to establish phonemes by solely observing the distributional properties of the input. A series of models has therefore considered whether this mechanism is appropriate for identifying individual phones. These models can be said to closely resemble the first stage of the two-stage approach that was described in Dillon, Dunbar, and Idsardi (2013). A further set of models which identify phonemic categories alongside allophonic rules, word forms and semantic topics have also been presented to determine the extent to which distributional learning has a mutually beneficial relationship with the identification of other levels of linguistic structure. By considering multiple mechanisms, these models provided explicit implementations of the single-stage model that was described in Dillon, Dunbar, and Idsardi (2013).

These two approaches were considered in a set of clustering models that were applied to distributions of F_1 and F_2 that were sampled from Inuktitut IDS (Dillon, Dunbar, and Idsardi, 2013). The vowel system of Inuktitut has three phonemic vowels, /i/, /a/, and /u/. Before uvulars, each of these vowels are realised as the allophones [e], [ɑ], and [o] respectively. These models indicated that the two-stage model was not viable as these models failed to identify these six clusters which corresponded to these allophonic units. Infinite Gaussian mixture models typically recovered three clusters with a pairwise F-score of .65 and thus collapsed the allophonic distinctions. When these models did recover six categories, the model outputs did not correspond well to the allophonic distinctions that differed in vowel height: pairs of categories often differed in backness rather than height, for example. Any errors that are introduced in the first stage of the two-stage approach are critical as they severely limit infants' abilities to identify a conditioning environment in the next stage.

An allophonic model which implemented the single-stage approach to perceptual attunement showed greater success (Dillon, Dunbar, and Idsardi, 2013). The Inuktitut vowel tokens that formed the input of this model were labeled with an indicator of whether the following consonant was uvular alongside measures of the first two formants. This model attempted identify the mean and variance of an unknown number of categories as well as a set of allophonic rule for each category. These rules established an association between pairs of categories: one that occurred before uvulars and another that occurred elsewhere. These paired categories had the same variance and the rule indicated the difference in their central tendencies. Models applied to the Inuktitut data successfully identified three pairs of categories with a pairwise F-score of .75. Since each category corresponded to a point vowel and the rules indicated that allophones had a more open quality, these results provided strong support the single-stage approach to perceptual attunement.

Improved model performance has also been observed in models which simultaneously learnt vowel categories and lexical items (Feldman et al., 2013; Frank, Feldman, and Goldwater, 2014). For each token, these models were presented with a measure of the first two formants and a lexical frame. These models attempted to learn the twelve vowel categories of American English presented in Hillenbrand et al. (1995) and a set of lexical items based on word frequencies in parental productions reported in CHILDES (Li and Shirai, 2000; MacWhinney, 2000). For example, a token with low formant values from the frame /b_k/ should be assigned to the category /ʊ/ and to the word form, ⟨book⟩. This lexical-distributional model consistently identified the twelve monophthongal vowel categories of American English with a pairwise F-score of .92 (Feldman et al., 2013). This improved performance can be contrasted with distributional models that recovered eight categories and had a pairwise F-score of .52. Lexical

set	N	categories
full	25	p, b, t, d, k, g, tʃ, dʒ, f, v, θ, ð, s, z, ʃ, ʒ, h, m, n, ŋ, l, r, w, j
place, manner	15	p/b, t/d, k/g, tʃ/dʒ, f/v, θ/ð, s/z, ʃ/ʒ, h, m, n, ŋ, l/r, w, j
manner only	6	p/b/t/d/k/g, tʃ/dʒ, f/v/θ/ð/s/z/ʃ/ʒ/h, m/n/ŋ, l/r, w/j

Table 9: The sets of consonants used in the word frames in Frank, Feldman, and Goldwater (2014). Symbols that are separated by slashes indicate distinctions which were collapsed when consonantal identity was ambiguous.

information supported the identification of phonetic categories each word type could be assumed to contain instances of the same vowel category. The identification of phonetic categories also supported the identification of lexical distinctions as the vowel in ⟨book⟩ had different acoustic properties from the vowels in ⟨back⟩ and ⟨bike⟩.

A further set of models have demonstrated that access to weak semantic cues also facilitates the identification of vowel categories (Frank, Feldman, and Goldwater, 2014). In this model, each vowel token was label with a situational context in addition to the set of formant measures and word frames seen in previous models. This model attempted to learn topics, word forms and vowel categories. For example, the word ⟨book⟩ could be associated with the activity of reading while the word ⟨bike⟩ be associated with the outdoors. Models that had access to semantic cues had significantly greater accuracy than models that only learnt word forms and phonetic categories. Situational information facilitated word learning as it ensured the robust identification of word forms that were frequent in a single context and allowed for similar word forms that did not share contexts to be distinguished. This study presented a further series of models that showed that lexical information facilitated phonetic category learning even when the relevant word forms presented ambiguous information about consonantal identity. Other models in this domain can be criticised for their ecological validity as the identity of consonants was fully specified in the word frames that were presented in the input (Feldman et al., 2013). This assumption does not align with the observation that infants show language-specific perceptual behaviours for vowels before they do so with consonants. Three different sets of consonantal categories were used in a further set of lexical-distributional models (Frank, Feldman, and Goldwater, 2014) and these are displayed in table 9. One set neutralised voicing distinctions and thus ⟨book⟩ shared a frame with words such as ⟨bug⟩ and ⟨poke⟩. The other set only provided the manner of each consonant and thus ⟨book⟩ shared a frame with ⟨kid⟩, ⟨tube⟩, and any other CVC forms with stop consonants. Though these models showed poorer performance than the fully specified models, models that were provided with word frames with only still resulted in improved performance in comparison to the models that exclusively learnt vowel categories from formant distributions.

2.2.3 Closing statements on distributional learning

Distributional learning in laboratory contexts has provided empirical support for a potential mechanism that explains perceptual development in infancy (Maye, Werker, and Gerken, 2002). This mechanism provides an explicit accounts of how the acoustic properties of the input shape infants' perceptual behaviours throughout the first year of life. The use of this mechanism requires the existence of a one-to-one relationship between native language categories and individual modes in the frequency distribution of the acoustic signal that infants are exposed to. Though experimental tasks present learners with low variance categories that are well dispersed in acoustic space, vowel categories in the input show a considerable degree of overlap. Computational models which have attempted to identify an unknown number of categories from a given sample of acoustic data have been used to replicate this learning task. These models have failed to identify phonemic units when they are presented with formant distributions that are representative of actual speech. This suggests that phonemic categories cannot be learnt through the observation of statistical regularities in the input and that infants must combine generalisations about the acoustic signal with their emergent knowledge of other levels of linguistic structure. Though distributional models have been applied to IDS and ADS data, studies in this domain have not drawn explicit comparisons of model performance across registers. If the properties of vowel production in IDS do promote the acquisition of linguistic structure, comparative analyses of these two registers may indicate that distributional learning can provide learners with a relevant set of native language categories.

2.3 Hyperarticulation in infant-directed speech

The second goal of this thesis is to address the extent to which vowel production in IDS facilitates the the identification of native language categories. Caregivers have been reported to modify their speech when addressing infants in most well-reported languages (see Cristia, 2013; Saint-Georges et al., 2013; and Soderstrom, 2007 for recent reviews). This specialised register features adaptations that have been claimed to facilitate the identification and processing of linguistic structure. The promotion of language acquisition provides a functionalist explanation for the observation that speech addressed to children bears the same properties across a large number of the world's languages. In terms of morphosyntax, IDS presents learners with shorter utterances, fewer disfluencies and a series of specialised lexical items. IDS forms such as *bunny, tummy* and *choo-choo* contrast with ADS *rabbit, stomach* and *train.* This register also features a slower speech rate, a global increase in pitch and larger pitch excursions. Infants preferentially attend to this register and this observation may partially explain the pedagogical features of this register (Cooper and Aslin, 1990; Fernald and Kuhl, 1987). Generalised effects of increased attention may be observed as learners presented with IDS input outperformed those presented with ADS in tasks involving word segmentation (Thiessen, Hill, and Saffran, 2005), word recognition (Singh et al., 2009; Song, Demuth, and Morgan, 2010) and the identification of phrasal boundaries (Kemler Nelson et al., 1989).

The hyperarticulation of vowels in IDS is another feature that has been claimed to promote language learning. The hyperarticulation hypothesis proposes that caregivers modify phonetic categories when they address infants in order to highlight native language distinctions distinctions (Bernstein Ratner, 1984; Kuhl et al., 1997). Descriptions of vowel production in IDS are relevant both to this hypothesis and to the viability of distributional learning in infancy. If this statistical mechanism plays a central role in perceptual attunement, speakers should be expected to modulate their speech such that categories show a lesser degree of overlap in acoustic space. Input of this type would provide learners with reliable statistical cues to the inventory of their native language. Modifications that facilitate learning will be referred to as cases of contrast enhancement while modifications that hinder the learner will be referred to as cases of contrast deterioration. The following section will review the phonetic studies that have aimed to address this hypothesis as well as methods that have been used to compare the relative discriminability of vowel distinctions in IDS and ADS.

2.3.1 Bernstein Ratner: an early view of IDS

Bernstein Ratner (1984) is an early study which considered the extent to which the properties of IDS vowel production facilitate the perception of native language distinctions. The acoustic analysis considered samples of naturalistic speech produced by nine female American English speakers. These caregivers were recorded in interactions with their own children and with an adult experimenter. These nine speakers were divided into three groups on the basis of their child's linguistic development as indicated in table 10. At the start of the study, three speakers had children who were preverbal, three had children in the holophrastic stage and three had children with a mean length of utterance (MLU) between 2 and 3.5. The IDS section of the corpus consisted of unstructured play sessions and the ADS section consisted of directed interviews with the adult experimenter. In these interviews, the adult experimenter attempted to elicit words that the speaker had uttered in the play sessions so that words appeared in both registers. The phonetic analysis presented in Bernstein Ratner (1984) only compared vowel tokens across registers if they were produced by the same speaker in same word type and syntactic context. Caregivers were recorded at eight week intervals in order to track how children's linguistic development affected vowel production. Children in the preverbal and holophrastic groups all advanced to higher developmental stages across these recording sessions.

The acoustic analysis of each register considered differences in the quality of nine monophthongal vowels (/i/, /ɪ/, /ɛ/, /æ/, /ɑ/, /ʌ/, /ɔ/, /u/, /ʊ/). This analysis considered the mean and standard deviation of the first two formants of each vowel in order to assess distributional properties of register. Specifically, this analysis adopted

Group	Infant Name	Infant Age 1	2	3
Prelinguistic	Amelia	1;6	1;8	1;10
	Dale	1;5	1;7	1;9
	Kay	1;1	1;4	1;6
Holophrastic	Alice	1;1	1;3	1;5
	Cindy	1;8	1;10	2;0
	Marie	1;6	1;8	1;10
Advanced	Anne	1;5	1;8	1;10
	Gail	1;9	1;11	2;1
	Lena	1;7	1;10	2;0

Table 10: Age of addressed infants in the corpus used in Bernstein Ratner (1984) across recording sessions, divided by their developmental level in the first recording session.

measures of precision and ambiguity. Vowels were considered to be more precise if they had a more peripheral quality, as indicated by Euclidean distance from the centre of the vowel space. Vowels were considered to be more ambiguous if they showed greater degree of overlap in acoustic space. The relative ambiguity of each register was defined by plotting a series of ellipses that indicated the variability of each category in the system. Each ellipse indicated an area that was one standard deviation away from the centre of each category. Cases of overlap were identified through a visual inspection of these vowel plots.

These measures indicated that the discriminability of vowels in IDS increased as a function of children's linguistic development. Speech to preverbal infants was found to be comparable to ADS both in terms of precision and ambiguity. Conversely, vowels in speech to the advanced children were more peripheral than ADS vowels and showed only a minimal degree of overlap. Speech addressed to children in the holophrastic stage was intermediate between the other two developmental groups.

Two follow-up analyses established that this effect of enhancement was independent of differences in word type and vowel duration across registers. Since function words undergo phonological reduction, greater peripherality in IDS may have been explained through the greater proportion of content words observed in this register. This explanation was ruled out since vowels in content and function words were equally peripheral in speech addressed to advanced children. Peripheralisation in IDS may also have been explained as a side-effect of greater vowel duration in IDS. Increased vowel duration provides speakers with more time to reach their intended articulatory targets and thus reduces undershoot. Register-specific differences in vowel production were not consistent with this explanation as IDS vowels did not have a greater duration than those in ADS.

These phonetic observations provided the initial evidence that the features of vowel production in IDS may facilitate the identification of native language vowel distinctions. The lesser degree of overlap observed in IDS was consistent with the idea that this register provides learners with reliable statistical cues to vowel identity. The implications of these results should be interpreted with caution, however. Both peripherality and overlap were based on subjective descriptions of the phonetic data. Thus, the observed effects may have just been tendential and therefore did not provide substantive empirical support for the hyperarticulation hypothesis. Statistically significant differences in objective measures of the distributional properties of each register are required to support or refute the existence of register-specific effects.

2.3.2 Vowel space expansion

Measures of the area of the vowel space were first presented in Kuhl et al. (1997) and the majority of the assessments of the hyperarticulation hypothesis have adopted this measure when assessing the relative discriminability of vowel distinctions in IDS and ADS. This measure is the area of the triangle defined by the mean values of the first two formants of the three point vowels /i/, /ɑ/, and /u/. These three vowels are found in the majority of the world's languages and represent the acoustic and articulatory extremes of vowel production (Ladefoged and Maddieson, 1996). The area of the vowel space is expected to be larger in IDS than ADS as this would indicate the central tendency of the three point vowels are more easily separated in this register. The observation of a larger vowel space implies greater dispersion and indicates that native language distinctions can be discriminated more easily.

Evidence of vowel space expansion in IDS was first observed in a study that considered the speech of American English, Russian, and Swedish mothers (Kuhl et al., 1997). For each language, ten caregivers were recorded speaking to their own infants who were aged between 0;2 and 0;5 as well as to an adult speaker of their language. For English and Russian, the phonetic analysis considered vowels produced in specific words that were elicited by asking the maternal speakers to discuss a series of toys. The analysis of Swedish vowels considered every instance of a point vowel that speakers produced. In each of these languages, caregivers produced a significantly larger vowel space in IDS relative to ADS. Other studies of English have observed vowel space expansion when American English caregivers addressed infants aged 0;4 and 0;11 (Cristia and Seidl, 2014) and when Australian English caregivers addressed infants aged 0;11 (Kalashnikova, Carignan, and Burnham, 2017). Similar cases of expansion were observed in Mandarin Chinese addressed to two groups of infants aged 0;6–0;8 and 0;10–1;0 (Liu, Kuhl, and Tsao, 2003) and in Japanese addressed to infants between 1;6 and 2;0 (Miyazawa et al., 2017).

Extensions of Kuhl et al. (1997) This measure of dispersion has been adopted to further consider how the properties of this register differ as a function of the developmental level or linguistic capabilities of the addressee. A longitudinal study of vowel production observed American English mothers when their infants were 0;11, 1;6 and 2;0 (Hartman, Bernstein Ratner, and Newman, 2016). Though IDS showed greater expansion than ADS at each point in development, the majority of speakers showed a trend where the degree of expansion decreased with infant age. This trend was in the opposite direction to that which was observed in Bernstein Ratner (1984) where the clearest vowel distinction occurred in speech addressed to infants aged 1;5 or older. No age-related effects on the degree of expansion were observed in a study of Mandarin Chinese speakers (Liu, Tsao, and Kuhl, 2009). A comparable degree of expansion was observed across speech infants aged 0;7–1;0 and that addressed to five-year-old children.

Comparisons between IDS and another registers have demonstrated that the linguistic capabilities of the addressee affect vowel production. An analysis of Australian English vowels compared speech addressed to infants aged 0;6 to speech addressed to a pet and ADS (Burnham, Kitamura, and Vollmer-Conna, 2002; Xu, Burnham, Kitamura, and Vollmer-Conna, 2013). Though IDS and pet-directed speech were found to be comparable in terms of affect and prosodic features, vowel space expansion was only observed in IDS. The area of the vowel space was comparable across pet-directed speech and ADS. These comparisons therefore indicated that expansion was not merely a feature of highly affective speech and suggested that IDS serves a pedagogical function. Measures of vowel space expansion have also been compared across IDS, ADS and foreigner-directed speech (Uther, Knoll, and Burnham, 2007). When compared to ADS, British English females produced an expanded vowel space when addressing infants aged between 0;4 and 1;0 and Chinese adults who spoke accented English. Since the degree of expansion was comparable across IDS and foreigner-directed speech, this suggested that speakers expanded their vowel space in order to accommodate listeners with a lesser degree of linguistic competence. This study also indicated that only IDS showed greater positive affect, indicating that existence of expansion is not a side-effect of the affective properties of this register.

Vowel space expansion has also been adopted in order to explore how the hearing status of infant addressee affects IDS vowel production. Such effects have been observed in Australian English-speaking mother of twins, one with normal hearing and the other who had hearing aids installed at 0;4 (Lam and Kitamura, 2010). Samples of IDS and ADS speech were collected when her children were aged 1;3 and 2;1. While vowel space expansion relative to ADS was observed in speech to the normal-hearing twin at both ages, speech to the hearing-impaired twin had a smaller vowel space than ADS. No effect of hearing status was found in an analysis of American English mothers of normal hearing and hearing-impaired children between 0;5 to 2;3 (Wieland et al., 2015). The degree of expansion in speech to hearing-impaired children was comparable to speech addressed to two groups of hearing children. These groups provided age matches for the hearing-impaired children in terms of chronological age and amount of hearing experience. Differences in the degree of expansion observed in IDS may instead be explained by infants' responsiveness (Lam and Kitamura, 2012). Australian English mothers were recorded interacting with infants aged 0;6–0;7 and adults through a video interface. This paradigm allowed for a simulated manipulation of infant hearing status by muting the audio. Speakers' beliefs about hearing status were also investigated by falsely informing the subjects that there was a fault with the audio. Vowel space expansion in IDS relative to ADS was observed when mothers had true beliefs that their infants could hear them. Beliefs about hearing status did not affect production as speakers produced an expanded vowel space even when they were falsely informed that their infants could not hear them. Infant feedback affected vowel production since a lack of expansion was only observed when the infant genuinely could not hear their mother. When interpreted alongside studies of speech to non-infant addressees, these results suggest that caregivers produce an expanded vowel space when addressing listeners that can benefit from the increase in discriminability that is associated with these modifications.

Criticisms of vowel space expansion The relevance of this measure has been brought into question since studies have failed to observe expansion and because this measure provides a limited view of the distributional properties of the input. Vowel space expansion is a universal property of IDS as studies that used the same methodology as Kuhl et al. (1997) did not replicate its effects. No evidence of expansion relative to ADS was observed in Danish addressed to children aged 1;7 (Bohn, 2013) or in Cantonese addressed to infants aged 0;3–1;0 (Xu Rattanasone, Burnham, and Reilly, 2013). Though the degree of expansion in Cantonese IDS did not differ across age groups, a series of individual comparisons revealed that the area of the vowel space in speech to infants aged 0;3 was smaller than that observed in ADS. Contraction of the vowel space has also observed in Dutch IDS addressed to infants aged between 0;11 and 1;4 (Benders, 2013). Further to this, American English caregivers produced comparable vowel space areas when they read a storybook to children aged between 0;3–1;8 and when they read to adults (Burnham et al., 2015). No difference in the area of the vowel space was observed across registers when French, British English and Japanese mothers read to children aged 0;6–1;10 and to adults (Dodane and Al-Tamimi, 2007). An analysis of IDS and ADS produced by six Norwegian mothers also no register-specific differences in the area of the vowel space (Englund and Behne, 2006). This analysis considered the speech that caregivers produced in free interactions with their infants from birth to the age of 0;6 and as well as similar interactions with other adults.

Results in this domain have assessed by considering the extent to which the area of the vowel space indicates the quality of the distributional information in the input. Vowel space expansion is often associated with contrast enhancement since this measure indicates that there is greater dispersion between the three point vowels. However, the quality of the distributional information in the input requires the consideration the central tendency, variance and frequency of each category in the system. Vowel space expansion is not sensitive to differences in the variance or frequency of categories across registers and only provides a composite measure of the dispersion for a subset of the vowels in the system. This approach should therefore be contrasted with the methods used in Bernstein Ratner (1984). Though the measures of peripherality and overlap were not objective or statistically verified, this analysis did address the dispersion and variance of a larger set of English vowel categories. Further to this, studies of adult production and perception have shown that the area of the vowel space is not a predictor of intelligibility (Neel, 2008). This analysis considered how the individual differences in the acoustic properties of American English vowels across a sample of speakers associated with intelligibility ratings for each of those speakers: this data was reported in Hillenbrand et al. (1995). The relative intelligibility of speakers did not correlate with the area of the vowel space that they produced. A closer examination revealed that listeners frequently misperceived two contrasts, /æ, ɛ/ and /ɑ, ʌ/. The dispersion of these specific vowels provided a stronger predictor of individual differences in intelligibility.

2.3.3 Other measures of vowel hyperarticulation

Since the area of the vowel space is not a sufficient indicator of the distributional properties of the input, comparative phonetic studies of IDS and ADS have adopted other measures in order to address the hyperarticulation hypothesis. Specifically, these measures have attempted capture the central tendency and variance of categories in acoustic space. These two measures have also been combined in order to estimate the degree of overlap between categories.

Central tendencies Two measures have been used to determine whether the alternations that caregivers make to the central tendency of vowel categories facilitate learning in infancy. Measures of peripherality consider the Euclidean distance of each category from the centre of the vowel space while dispersion is defined as the Euclidean distance between pairs of categories. An analysis of eight vowels of American English found that not all vowels were more peripheral in IDS (McMurray et al., 2013). This analysis considered vowel production that was elicited by asking caregivers to read storybooks which featured the target words that are presented in table 11 to their infants aged 0;9–1;1 and to an adult experimenter. Though /oʊ/ and /æ/ were found to be more peripheral in IDS, /ɝ/ and /ʌ/ did not differ in peripherality across registers and /ɑɹ/ was centralised in IDS. Though centralisation may be interpreted as evidence of hypoarticulation, this pattern did not necessarily indicate a case of contrast deterioration. Under the principle of maximal dispersion, categories are optimally separated when larger vowel systems consist of both peripheral and central vowels (Liljencrants and Lindblom, 1972; Lindblom, 1986). This analysis of American English did not report the Euclidean distance between categories in order to determine whether the simultaneous occurrence of peripheralisation and centralisation resulted in greater separation in IDS (McMurray et al., 2013).

Other studies have considered dispersion as a direct indicator of the discriminability of vowel distinctions of the input. Measures of dispersion were applied to two tense-lax

vowel target	minimal pair
[iɹ]	deer, tear
[eɹ]	bears, pears
[æ]	baths, baths
	gaps, caps
[ɝ]	girls, curls
[ʌ]	bugs, pugs
[ɑɹ]	darts, tarts
	guards, cards
[ɑɪ]	dime, time
[oʊ]	goats, coats
	bowls, poles

Table 11: The set of words that were used to elicit IDS and ADS vowels in McMurray et al. (2013).

contrasts, /i, ɪ/ and /eɪ, ɛ/, produced by American English mothers when addressing infants aged 0;4 and 0;11 and adults (Cristia and Seidl, 2014). The acoustic distinction between these vowels was operationalised as the log duration of each vowel and the value of F_1 and F_2 taken at 40% and 80% of the vowel's duration. Though both sets of speakers produced an expanded vowel space in IDS, neither set of speakers showed significant increase in dispersion in IDS for either of the distinctions. A comparison of the IDS and ADS productions of Danish mothers revealed no increase in dispersion when they spoke to children aged 1;7–1;8 (Bohn, 2013). This analysis indicated that the area of the vowel space and the distance between each point vowel were comparable across registers. No effect of dispersion was observed when the dispersion of three further distinctions were compared across registers (/i, e/, /eː, ɛː/, /oː, ɔː/). The only significant difference that was observed across registers indicated that /eː/ and /ɛː/ were closer in IDS than ADS, highlighting a case of deterioration.

Within-category variance and overlap The quality of distributional information depends on measures of within-category variance. As highlighted in Cristia and Seidl (2014), greater dispersion in IDS will only result in a lesser degree of overlap if within-category variance remains comparable across registers. The degree of overlap must therefore be compared across registers, especially since IDS vowel production has been reported to be more variable than ADS across a number of languages (American English, Russian and Swedish: Kuhl et al., 1997; American English: Cristia and Seidl, 2014, Kirchhoff and Schimmel, 2005, McMurray et al., 2013; Dutch: Benders, 2013; Japanese: Miyazawa et al., 2017). Studies that have compared the degree of overlap in IDS and ADS have not provided evidence of contrast enhancement in IDS. The comparative analysis of the American English distinctions /i, ɪ/ and /ɛ, eɪ/ also considered measures of overlap alongside the measures of dispersion discussed above (Cristia and Seidl, 2014). The degree of overlap between /i/ and /ɪ/ was comparable across registers while /ɛ/ and /eɪ/ showed greater overlap in IDS. This lack of enhancement followed from the observation of greater variance and comparable dispersion in IDS. A comparative analysis of vowel production in Japanese failed to demonstrate an effect of enhancement on the basis of measures of overlap (Miyazawa et al., 2017). This study considered speech from the RIKEN Japanese Mother-Infant Conversation Corpus (R-JMICC: Mazuka, Igarashi, and Nishikawa, 2006) which details a caregivers' interactions with infants aged between 1;6 and 2;0 and with adult speakers. The IDS and ADS productions of each speaker were further compared to their production of careful read speech. This analysis applied measures of overlap to ten vowel distinctions by pairing each of the five short vowels of Japanese /i/, /ɛ/, /a/, /o/, and /ɯ/). The degree of overlap was comparable across registers in IDS and ADS, indicating a lack of contrast enhancement in IDS. Comparisons with read speech provided further evidence against the hyperarticulation hypothesis. Vowels in this sample showed significantly less overlap than those in IDS and ADS.

2.3.4 Consonantal distinctions in IDS

Comparative analysis of consonants in IDS and ADS have also provided evidence against the hyperarticulation hypothesis. Analyses of this type have primarily considered the realisation of stop voicing. In IDS addressed to infants aged 0;3, Swedish caregivers produced a shorter VOT for voiced and voiceless stops in comparison to ADS (Sundberg and Lacerda, 1999). Analyses of stop voicing distinctions produced by Norwegian and American English caregivers have demonstrated the opposite effect (Englund, 2005; McMurray et al., 2013). Speakers produced longer VOT for both stops in IDS in both languages. The modifications that speakers made in IDS did not result in greater category dispersion and thus were not consistent with the hyperarticulation hypothesis. Contrast enhancement would require a reduction of the VOT of voiced stops, an increase in the VOT of voiceless stops, or a combination of both.

Computational models have been used to assess the relative discriminability of all of the distinctions in Japanese IDS and ADS (Martin et al., 2015). This analysis considered the degree of overlap in both consonant and vocalic distinctions in the IDS and ADS production data taken from R-JMICC (Mazuka, Igarashi, and Nishikawa, 2006). These segmental distinctions were assessed by pairing every possible syllable of Japanese. Each distinction therefore depended on comparisons of multiple syllable pairs. The distinction between /m/ and /n/ depended on pairs such as /mi, ni/ and /mɛ, nɛ/ the distinction between /i/ and /ɛ depended on pairs such as /mi, mɛ/ and /ki, kɛ/. The acoustic properties of each syllable were operationalised with mel-frequency filter banks that captured the spectral envelop of these units. The degree of overlap between these segmental distinctions was assessed through a simulated ABX task. This task presented the model with two tokens from different categories, A and B, and asked it to label a third, X, as an instance of either A or B. This model selected a response on the basis of the acoustic similarity of X to A and B. Accurate discrimination indicated that A and B minimally overlap while poorer performance indicated a greater degree of ambiguity. Comparisons of model performance indicated that there was small but significant effect of deterioration in IDS since these models showed poorer performance than the ADS models. Cases of contrast deterioration were observed for both vocalic and consonantal distinctions in caregivers' IDS productions.

2.3.5 Multidimensional acoustic data

This discussion presented an evaluation of different measures of discriminability and their relevance to the hyperarticulation hypothesis. However, methodological discussions of this type have rarely considered the acoustic dimensions that these measures are applied to. The majority of comparative analyses of vowel quality have operationalised these distinctions through measures of the first two formants. Though F_1 and F_2 are correlates of vowel height and vowel backness respectively, these are not the only acoustic dimensions that are relevant to vowel quality in American English. Analyses of the perception and production of vowel quality in American English have indicated that vowel duration, F3 and patterns of spectral change all contribute to these distinctions (Hillenbrand et al., 1995; Hillenbrand, 2013). Two discussions of vowel distinctions in IDS have advocated for the use of these additional dimensions when assessing the statistical properties of the input.

On the one hand, multidimensional approaches to vowel quality have been proposed as one way of mitigating the cases of overlap that have been observed in formant analyses of IDS vowel production (Swingley, 2009). The ambiguity observed in formant distributions is problematic since it hinders the use of distributional learning in infancy. If a broader set of acoustic dimensions provides the infant learner with relevant information about native language distinctions, this may allow for more robust statistical regularities to be observed in the input. In this way, analyses of IDS that have only considered the first two formants may overstate the difficulty of the learning task. Though this argument is framed in a discussion of the properties of IDS, it does not make explicit references to register-specific differences in the quality of distributional information. Instead, this facilitative effect of multidimensional information focusses on the absolute discriminability of vowels in a single sample of speech and is equally valid for analyses of both IDS and ADS.

This claim should be contrasted with the claim that multidimensional analyses of IDS and ADS provide a stronger evidence of contrast enhancement in IDS (Eaves Jr. et al., 2016). This study used computational models to identify a set of modifications to acoustic properties of American English vowels that were consistent with the hyperarticulation hypothesis. Specifically, these models made alterations to the first three formants of American English vowels that were reported in Hillenbrand et al. (1995). This process can be viewed as the inverse of the Gaussian mixture models that were discussed in 2.3.2. Rather than attempting to locate a set of categories that optimally fit a given set of data, this model aimed to locate a set of data that optimally predicted a given set of vowel categories. This simulated vowel data aligned with the principle of maximal dispersion as vowels were more peripheral and more central in the optimised data in comparison to the ADS data. Despite this, the optimised data did not show greater category dispersion two-dimensional formant space than the original dataset. Greater dispersion was only observed once the third formant was included. Under the assumption that caregivers similarly optimise IDS vowel distinctions in high dimensional space, these model results support the claim that multidimensional analyses may be required in order to identify patterns of enhancement in this register. It follows from this claim that formant analyses of IDS which have failed to demonstrate an effect of enhancement may have presented similar false negatives. Additionally, it is important to note that the variance of the optimised teaching data was greater than that of the original ADS data. Though greater variance in IDS has been associated with a greater degree of overlap, increased variance was viewed as a potentially facilitative property of the input in the current study. As indicated in figure 4, it has been proposed that high-variance categories may be distinguished easily if they also differ in orientation. These distinctions may be detected by learners as each category overlaps through each other, forming a conspicuous X-shaped configuration.

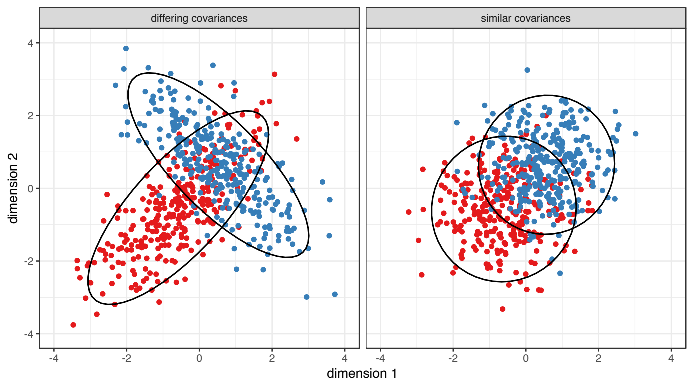

Figure 4: Distinctions which illustrate the facilitative effect associated with orientation in Eaves Jr. et al. (2016). The difference in orientation between the lefthand classes make the distinction easier to identify than that on the right.

Studies that have considered acoustic properties beyond F_1 and F_2 (Cristia and Seidl, 2014; Martin et al., 2015) have partially addressed each of these claims. The fact that neither of these studies observed an effect of hyperarticulation in IDS in spite of the use of high dimensional acoustic data stands against the claim made in Eaves Jr. et al. (2016). Neither of these studies drew explicit comparisons between their high dimensional analysis and formant analyses. However, Cristia and Seidl (2014) did further report that a lack of enhancement was also observed when acoustic dimensions were separately. Since both studies used the same corpus, the results of Martin et al. (2015) can be compared with a formant analysis that also found a lack of enhancement which was presented in Miyazawa et al. (2017). The claims made in Swingley (2009) cannot be assessed on the basis of these studies since these analyses focussed on the relative discriminability of IDS and ADS. In order to test these claims, it would be necessary to observe the effect that multidimensional data has on the absolute discriminability of a single register.

2.3.6 Other accounts of IDS vowel production

A major goal of this thesis is to assess the extent to which the properties of vowel production in IDS conform with the predictions of the hyperarticulation hypothesis (Bernstein Ratner, 1984; Kuhl et al., 1997). This hypothesis can be viewed a functionalist explanation of the phonetic modifications that caregivers produce when addressing infants and their ubiquity across the world's languages. The limited support for the hyperarticulation hypothesis has stimulated alternative accounts of the phonetic properties of this register. A consideration of these accounts may provide further insights the properties of IDS and could facilitate the interpretations of the results of the current analysis. However, this thesis does not intend to assess whether these alternative accounts provide a more apt description of the features of IDS than the hyperarticulation hypothesis does.

Prosody as an explanatory factor One account states that the register-specific properties that have been associated with hyperarticulation are a side-effect of the prosodic features of this register (McMurray et al., 2013). IDS has a slower speech rate, shorter utterances and larger pitch excursions than ADS (Fernald et al., 1989). Each of these factors reduce the likelihood that vowels in IDS will be phonetically reduced. Under this account, comparative analyses of IDS and ADS have overstated apparent effects of hyperarticulation because they have failed to adequately control for these differences across registers. The elicitation method used in Kuhl et al. (1997) oversamples prominent vowels since it focusses on those that occur in a small set of content words that refer to discourse-salient objects. Studies that have controlled for prosodic effects have shown minimal register-specific differences. For example, the VOT of American English stops did not differ across registers once prosodic factors had been controlled for (McMurray et al., 2013). Similarly, American English vowels in IDS and ADS showed differences in peripherality due to utterance position and syllable stress rather than register differences (Wang, Seidl, and Cristia, 2015). Analyses of how prosodic prominence affected the quality of nine American English monophthongs in IDS have been forwarded as evidence against this claim (/i, ɪ, ɛ, ʤ˞, ɑ, ʌ, ɔ, ʊ, u/: Adriaans and Swingley, 2012, 2017). Prosodically prominent vowels in IDS were found to be more peripheral than vowels in other positions. Because of this, these studies have been interpreted as evidence that register-specific differences in discriminability cannot be dismissed as the result of differences in the prosodic structure of IDS and ADS. These studies did not draw a comparison with vowel production in ADS. Therefore, it is unclear whether IDS showed greater peripherality than ADS or whether these effects of prominence were larger in IDS than ADS.

Affect as an explanatory factor Another alternative proposal has stated that the properties of IDS vowel production are consistent with greater positive affect, rather than any explicitly pedagogical effect (Benders, 2013). Positive affect in speech is expressed by raising the frequency of the first three formants. Following the frequency-size relationship, high frequencies imply a small body size and thus a lower threat level (Ohala, 1980, 1984). A study of the realisation of Dutch /i/, /aː/, /ɑ/, and /u/ demonstrated that each vowel had a higher F_2 and F_3 in IDS compared to ADS. Other phonetic studies reported results that were consistent with formant raising (greater F_1, F_2 for Australian English /i/ and /ɑ/: Burnham, Kitamura, and Vollmer-Conna, 2002; Kalashnikova, Carignan, and Burnham, 2017; Xu et al., 2013; greater F_1, F_2 for Norwegian /u/ and /ɑ/: Englund and Behne, 2006). Patterns where high vowels are more open and back vowels are more advanced in IDS were inconsistent with contrast enhancement. Further to this, the articulatory properties of Australian English IDS resembled those of highly emotional speech rather than being consistent with contrast enhancement (Kalashnikova, Carignan, and Burnham, 2017). Though an acoustic analysis indicated vowel space expansion in IDS relative to ADS, an articulatory analysis indicated that the position of a speaker's tongue and lips for /i/, /ɑ/ or /u/ did not differ across these two registers. These acoustic differences instead originated from the height of a speaker's larynx in IDS. Measures that approximated the length of the vocal tract showed similar results across IDS and highly emotional speech with each of these registers being distinct from ADS.

Both of these approaches refute the hyperarticulation hypothesis, stating that the modifications that caregivers make addressing infants are not motivated by the clarification of native language distinctions. However, it should be noted that refutations of this hypothesis are not incompatible with the existence of other facilitative effects in IDS. After all, the prosodic and affective properties of this register have been linked to the observation that infants preferentially attend to this register (Fernald and Kuhl, 1987). This ability to capture and hold infant attention may have a broad facilitative effect on the identification and processing of linguistic units that is independent of hyperarticulation. These types of effects have been demonstrated with regard to the segmentation (Thiessen, Hill, and Saffran, 2005) and recognition of word forms (Singh et al., 2009; Song, Demuth, and Morgan, 2010). Infants' emergent knowledge of the lexicon may therefore have an indirect effect on perceptual attunement since this process depends, in part, on interactions between the properties of the phonetic signal and other levels of linguistic structure.

2.3.7 Closing comments on the hyperarticulation hypothesis

The hyperarticulation hypothesis (Bernstein Ratner, 1984; Kuhl et al., 1997) forms part of a larger claim that the adjustments that caregivers make to their speech when addressing infants help to promote language acquisition. These facilitative effects provide an functionalist explanation for the observation that IDS bears the same properties across the majority of well-documented languages. This facilitative effect has primarily been observed through measures of the area of the vowel space. Expansion in IDS indicates that speakers make appropriate adjustments to the central tendency of a subset of vowels as defined by their first two formants. However, expansion has not always been observed in this register. Additionally, this measure provides only a partial view of the distributional properties of the input. If the hyperarticulation hypothesis is to be linked with the use of distributional learning in infancy, analyses of vowel production in IDS must consider differences in the central tendency, variance and frequency of categories across registers. Studies which have compared of the degree of overlap between vowel categories across registers have found that IDS vowels have a similar or greater degree of overlap in comparison to their ADS counterparts. Though this lack of a register-specific effect of enhancement challenges the hyperarticulation hypothesis, further study is required in this domain as variance-sensitive measures have only been reported for a small number of vowel distinctions. In order to identify the intentions behind the modifications that caregivers make in IDS, these measures must be applied to a broader set of distinctions in high dimensional acoustic space.

2.4 Chapter summary

In summary, the current chapter has introduced and motivated my two primary research questions: the first concerns the extent to which the properties of IDS vowel production facilitate the recognition of native language distinctions in infancy while the second considers the extent to which distributional learning can explain this learning task. Though perceptual attunement has typically been explained through a process of statistical inference over hyperarticulated input, the current chapter raises concerns about current characterisations of IDS and the mechanisms behind this learning process. With regard to the hyperarticulation hypothesis, the current chapter argues that there is limited evidence for the claim that caregivers modify their vowels when they address children in order to ensure that native language distinctions show a minimal degree of overlap. I will therefore present a novel analysis of a large, naturalistic corpus of American English IDS and ADS and apply multiple of measures of discriminability to a broader set of categories and acoustic dimensions. With regards to my second research question, the current chapter drew comparisons between the experimental evidence of the availability of distributional learning in infancy with computational models that attempt to replicate this task. Successful cases where infants that were exposed to distributional information showed an alteration in perceptual behaviour can be critiqued for presenting infants with idealised information concerning a single distinction. By contrast, the poorer performance of computational models can be associated with the fact that these models have learn entire inventories from naturalistic data where distributional information is ambiguous. To extend current assessments of the use of this mechanism in infancy, I will apply a series of clustering models and logistic regressions to multidimensional data that is sampled from both IDS and ADS.

3 Acoustic analysis of F_1 and F_2 in IDS & ADS

The comparative acoustic analyses presented in this thesis will be divided across the current chapter and the one that follows. Both chapters have a similar structure in that they present empirical data concerning differences in vowel production across IDS and ADS. They have a singular goal of assessing the extent to which acoustic data that strongly resembles in the input which infants receive is adapted to facilitate the identification and processing of American English vowel categories in infancy (Bernstein Ratner, 1984; Kuhl et al., 1997). This acoustic analysis will also enable a discussion of the viability of distributional learning in infancy. The properties of IDS and ADS will be compared objectively by applying a series of measures of discriminability to data sampled from each registers. Specifically, these measures consider the central tendency and their variance of vowel categories in caregivers' speech. Measures of the central tendency locate each category in acoustic space and while measures of variance describe their limits. These two chapters differ in that they consider a different set of acoustic dimensions to capture register-specific differences in vowel quality. The current chapter exclusively considers the value of the first two formants, allowing for direct comparisons with prior phonetic analyses that have adopted measures of the area of the vowel space and other measures of central tendency. The following chapter will consider a broader set of acoustic dimensions: measures of the third formant, vowel duration and patterns of spectral change will be considered individually as well as in combination with F_1 and F_2. This analysis will be compared with the formant analysis in order to determine whether multidimensional data provides stronger evidence of hyperarticulation in IDS and whether these additional dimensions mitigate the ambiguity that is apparent the input.

The current chapter extends previous investigations of the hyperarticulation hypothesis by supplementing comparisons of the central tendency of vowel categories in IDS and ADS with measures of variance and overlap. The current acoustic analysis will exhaustively consider the full set of American English vowels and the distinctions between each of them. In doing so, it will provide a fuller account of IDS vowel production than studies which have solely considered the area of the vowel space as an index of hyperarticulation.

3.1 Methodology

This section will describe the selection of a speech corpus which contains comparable samples of IDS and ADS and the partially automated acoustic analysis which considered the vowel tokens which occurred in the corpus data. Further to this, this section will outline the acoustic measures which were adopted to describe differences in vowel quality across registers as well as the measures of discriminability which indicated the extent to which IDS vowels were hyperarticulated relative to their ADS counterparts.

3.1.1 Materials

The current comparative acoustic analysis considered a subset of the speech corpus which was originally collected in Bernstein Ratner (1984). This subset of the corpus consists of a series of interactions between four mothers and their infants who were aged between 1;1 and 2;0. This subset was selected as it represented the set of speakers for whom acoustic data was accessible through the CHILDES database (MacWhinney, 2000). These four speakers will be referred to throughout the thesis using aliases which were derived from the pseudonyms that were given to their children in the original study. As indicated in table 12, speakers ALI and CIN had children who were in the holophrastic stage at the beginning of recordings while the speakers ANN and GAI had children who were in the advanced group with an MLU between 2 and 3.5. The IDS section of the corpus consisted of a series of unstructured play sessions lasting approximately twenty minutes while the ADS section of the corpus consisted of a series of directed interviews that were led by an adult researcher. Each mother-infant dyad was recorded in three separate sessions which were collected at eight-week intervals.

		Session		
Group	Alias	1	2	3
Holophrastic	ALI	1;1	1;3	1;5
Advanced	ANN	–	1;8	1;10
Holophrastic	CIN	1;8	1;10	2;0
Advanced	GAI	1;9	1;11	2;1

Table 12: The subset of the sessions originally recorded in Bernstein Ratner (1984) that had audio data for both IDS and ADS which was accessible through CHILDES (MacWhinney, 2000). Caregivers will be referred to using aliases based on their child's pseudonym.

3.1.2 Data extraction and acoustic analyses

The current acoustic analysis considered register-specific differences in the acoustic properties of fifteen American English vowel categories. This set of vowels comprised the twelve vowels which were analysed in Hillenbrand et al. (1995) (/i/, /ɪ/, /eɪ/, /ɛ/, /æ/, /ɝ/, /ʌ/, /ɑ/, /ɔ/, /oʊ/, /ʊ/, /u/) as well as three diphthongs (/aɪ/, /aʊ/, /ɔɪ/). These categories were uniquely paired to form a set of 105 distinctions.

The acoustic analysis that is presented in this chapter was partially automated through the use of the FAVE suite (Forced Alignment and Vowel Extraction; Rosenfelder, Fruehwald, Evanini, and Yuan, 2011). This suite consists of two tools that can be applied to large speech corpora. The first of these, `FAVE-align`, is a forced aligner which locates boundaries at the level of the word and the segment within samples of audio. The second tool, `FAVE-extract`, conducts a consistent, automated formant analysis which extracts acoustic data from the vowel tokens that are located using `FAVE-align`. Though acoustic analyses which make use of forced alignment and automated vowel extraction are common in the domain of sociophonetics (Eckert and Labov, 2017; Hall-Lew, Eiswirth, Valentinsson, and Cotter, 2017; Sonderegger, Bane, and Graff, 2017; inter alia), this methodology has not been widely adopted in comparative analyses of vowel production in IDS and ADS. Though forced alignment has been used locate vowel tokens in samples of IDS (Elsner and Ito, 2017; Kirchhoff and Schimmel, 2005; Ko and Soderstrom, 2013), no studies to my knowledge have used automated vowel extraction as a method of identifying differences in vowel quality across registers.

This discussion will now outline how this suite was used to facilitate an analysis of the IDS and ADS speech corpus that was described in 3.1.1. Except where specifically noted, this section will describe the default implementation of the FAVE suite (see Fruehwald, 2013; Labov, Rosenfelder, and Fruehwald, 2013 for further descriptions of this method). The current section will additionally outline the exclusion criteria by which certain vowel tokens that occurred in the corpus were removed from the final analysis.

Forced alignment In order to locate word and segmental boundaries, `FAVE-align` requires three inputs: an audio recording, a transcript of each utterance that occurs in the recording, and a dictionary of segmental transcriptions for each word that occurs in the transcript. Since `FAVE-align` and other forced aligners locate boundaries by incrementally processing the audio, this analysis requires transcriptions that capture every phonetic event that happens within the bounds of each utterance. Without this kind of transcript, this automated method may falsely identify these events as vowel tokens. I used the original transcriptions from Bernstein Ratner (1984) as a guide when transcribing each of the utterances that the speakers produced in both IDS and ADS. As well as documenting the words that speakers uttered, I additionally transcribed each filled pause, false start, and mispronunciation that was produced and indicated the start and end of these utterances. The transcriptions also detailed non-speech vocalisations such as laughter, gasps and coughs as well as any non-speech sounds that occurred within the bounds of an utterance. By default, `FAVE-align` uses the Carnegie Mellon University Pronouncing Dictionary to convert words in the transcriptions in a sequence of segments. I supplemented this resource with a series of additional segmental transcriptions for words that were specific to IDS or to individual speakers. `FAVE-align` combines the segmental and utterance-level transcripts into a sequence of segments which can be aligned with the audio.

When provided with these three types of input, the forced aligner uses a Hidden Markov Model to identity where the segmental boundaries occur in each utterance in the audio data that it is provided with. The forced aligner in the FAVE suite is built on `p2fa` (the Penn Phonetics Lab Forced Aligner: Yuan and Liberman, 2008, 2011). The Hidden Markov Model in `p2fa` was trained on labelled data sampled from 25.5 hours of audio data which consists of the oral arguments of eight speakers in the Supreme Court of the United States (SCOTUS). It learnt the identity of monophones which correspond to the segments of American English. This approach models the sequence of segments in each utterance as a series of hidden states. This process locates each hidden state by converting the audio data into a series of 10ms slices. The acoustic property of each slice are captured using a variant of cepstral coefficients. The model compares the acoustic quality of each slice sequentially to locate transitions between segments in the sequence. Since this model can only insert a boundary between slices, the resolution of this process is fixed at 10ms. I manually inspected each boundary that was place in the forced alignment to identify errors in this automated process. Boundaries which were not located within 10ms of the genuine onset or offset of the vowel were corrected by hand as well cases where the identity of vowels were not correctly identified. The latter occurred when the Hidden Markov Model failed to reduced forms (e.g. *that* can be realised as [ðæt] or [ðət]) or forms with alternate pronunciations (e.g. *either* can be realised as [aɪðɝ] or [iːðɝ]). These verified and corrected vowel tokens served as the input to `FAVE-extract`.

Vowel extraction In order to assess register-specific differences in the quality of vowels in caregivers' productions in IDS and ADS, I used `FAVE-extract` to automate the process of estimating the value of the first two formants of each vowel token. This tool automates formant analyses by using Bayesian inference to select a single optimal analysis from a series of linear predictive coding (LPC) analyses with different parameter settings. For each relevant vowel token, `FAVE-extract` generates four LPC analyses which respectively attempt to locate 3, 4, 5, and 6 formants in a range of frequencies between 0 and 5500Hz. Different parameter settings are required for vowels of different qualities. Analyses which locate a smaller number of formants may show better performance for vowels that have well separated formants such as /i/, where F_1 is low and F_2 is high. Analyses with a larger number of formants are disfavoured since they may spuriously locate an additional formant between the first and second formants. Conversely, analyses which locate a greater number of formants may show better performance for vowels with similar formants such as /u/, where both F_1 and F_2 are low. Analyses with fewer formants are disfavoured since they may erroneously merge the first and second formants. An optimal analysis is selected from amongst the four candidates by comparing each of them to a distribution of expected values for the relevant vowel categories. These distributions of expected values are known as priors and consist of values for the mean and bandwidth of each formant that are reported in the Atlas of North American English (Labov, Ash, and Boberg, 2005). The similarity of each candidate to the prior for the relevant category is calculated using Mahalanobis distances. The selection of an optimal analysis can be viewed as a process of Bayesian inference as the priors which the candidates are compared to are updated through an iterative process. Each subsequent run uses the values for the mean and bandwidth of the first two formants that were selected as optimal in the previous run as a prior. This process iterates until the model consistently selects the same set of candidate analyses as optimal. Iterating the process ensures that the formant analysis is representative of each speaker's productions rather than the reference data.

Exclusion criteria This process of automated alignment and extraction had the potential to identify a total of 27,065 vowel tokens from the speech corpus as a whole. This absolute maximum number of tokens represents the productions of four speakers, consisting of 15,687 IDS vowel tokens and 11,378 ADS vowel tokens. However, I excluded many of these tokens from the final analysis on the basis of three broad criteria which are indicated below. The exact number of the tokens that were excluded from the analysis are provided in table 13 for the IDS section of the corpus and in table 14 for the ADS section of the corpus.

IDS

criterion	ALI	ANN	CIN	GAI	total
initial	3506	2366	5290	4525	15687
alignment	249	182	542	514	1487
<50ms	565	588	900	1000	3053
schwas	276	93	329	163	861
outliers	487	309	688	568	2052
total	1929	1194	2831	2280	8234

Table 13: Details of the vowel tokens which were excluded from the IDS corpus data.

ADS

criterion	ALI	ANN	CIN	GAI	total
initial	2479	2551	3282	3066	11378
alignment	112	322	239	196	869
<50ms	487	575	982	836	2880
schwas	221	166	171	151	709
outliers	325	284	412	448	1469
total	1334	1204	1478	1435	5451

Table 14: Details of the vowel tokens which were excluded from the ADS corpus data.

(1) Tokens were excluded if they could not be reliably identified as an instance of one of the fifteen relevant categories

(2) Tokens were excluded if their acoustic properties did not allow for measures of duration and formant values to be taken.

(3) Tokens were excluded if the automated vowel extraction process gave anomalous results.

Tokens were first excluded from the analysis during the manual inspection of the forced alignment that `FAVE-extract` generated. Tables 15 and 16 provide further detail concerning vowel tokens that were excluded during this stage of the analysis pipeline. Vowel tokens which overlapped with background noise or other speech were excluded from the analysis as their acoustic properties could not be isolated from these other sounds. The identity of vowel tokens that occurred in noise could not be determined reliably in some cases. Because of this, sections of audio which was excluded on the basis of these two criteria may have contained more than one vowel token. The number of excluded tokens for these criteria thus represent minimum values rather than absolute counts. Cases of disfluent speech, such as false starts and mispronunciations, were excluded since they did not represent canonical realisations of a specific category. Elided vowels and sung speech was also excluded for not being canonical tokens. The requirement that vowel tokens were suitable for the tracking of formant values across their entire realisation resulted in a further set of exclusions. Instances of laughter and creaky voice were excluded as these vowel tokens did not have continuous periodic energy throughout their duration. Vowels with a low amplitude or with a lack of periodic energy, such as whispered speech, were also excluded for this reason. All of the vowel tokens which remained after these exclusions were analysed acoustically through the use of `FAVE-extract`.

IDS

criterion	ALI	ANN	CIN	GAI	total
background noise	26	10	83	79	198
speaker overlap	44	43	61	40	188
disfluencies	25	26	77	54	182
elided vowel	38	50	142	123	353
singing	7	5	5	44	61
creaky voice	11	9	8	12	40
laughter	5	4	17	2	28
low amplitude	39	11	42	97	189
whispering	54	24	107	63	248
total	249	182	542	514	1487

Table 15: Details of the vowel tokens which were excluded during the transcription and alignment of the IDS corpus data.

Tables 13 and 14 indicate that vowel tokens were also excluded during or after the process of extracting acoustic dimensions from the corpus data. By default, `FAVE-extract` does not analyse vowels with a duration of 50ms or less and thus tokens of this duration were excluded from the analysis. Any vowels that were labelled as instances of [ə] in the segmentation were excluded from the analysis since only categories that are phonemic in American English were considered to be relevant. Finally, the mean value and variance of F_1 and F_2 for each vowel category were considered in order to identify outliers within the productions of the four speakers. These measures

ADS					
criterion	ALI	ANN	CIN	GAI	total
background noise	13	13	21	36	83
speaker overlap	45	228	102	45	420
disfluencies	6	17	42	10	75
elided vowel	26	30	36	43	135
laughter	0	9	24	1	34
singing	0	0	0	0	0
creaky voice	5	2	0	3	10
low amplitude	13	12	7	52	84
whispering	4	11	7	6	28
total	112	322	239	196	869

Table 16: Details of the vowel tokens which were excluded during the transcription and alignment of the ADS corpus data.

were used to calculate the Mahalanobis distance of each vowel token relative to the category which it was a member of. This measure indicates the distance between a data point the centre of a category or distribution in terms of the standard deviation of that category. Tokens with a Mahalanobis distance of 3 or greater were excluded as outliers. A further set of outliers were identified by visually inspecting the data: tokens of /ɑ/ which had an F_1 that was less than or equal to 500Hz were excluded manually.

3.1.3 Measures of discriminability

This comparative formant analysis of IDS and ADS presented measures of the area of the vowel space to allow for direct comparison with previous studies which have adopted this measure as an index of hyperarticulation in IDS. This thesis adopted two further measures of discriminability that considered differences in the central tendency of F_1 and F_2 across registers. The distance between each category's central tendency and the centre of the vowel space captured differences in peripherality across registers while distances between the central tendency of paired categories capture differences in dispersion. This chapter also considered global differences in the value of the first two formants across registers since raised formant values have been proposed as an indicator of positive affect in caregivers' speech (Benders, 2013). Two further statistics considered how the variability of caregivers' productions across registers affected the distributional properties of the input. Within-category variance was measured directly as the standard deviation of F_1 and F_2 for each vowel category while D(a) (Newman, Clouse, and Burnham, 2001) captured differences in the degree of overlap between paired categories across registers. This chapter also presented measures of S2 (Garcia, 2012), an indicator of the orientation of paired categories, in order to evaluate the claim that speakers may enhance distinctions in IDS by adjusting the covariance of categories in this register.

Though previous analyses have indicated that the frequency of individual categories affects the quality of distributional properties of the input Bion et al., 2013, none of the statistics which were adopted in this chapter were sensitive to differences in the frequency of vowel categories. Because of this, it is possible that these measures overstate the discriminability of low frequency categories in the input. The frequency of each category across speakers and registers is reported in table 17.

	ALI		ANN		CIN		GAI	
	IDS	ADS	IDS	ADS	IDS	ADS	IDS	ADS
i	10.6	13.9	11.4	14.4	11.0	13.1	13.0	13.4
ɪ	13.2	10.5	9.9	9.1	11.5	10.1	12.7	9.7
eɪ	4.7	4.6	8.0	4.4	6.2	6.3	6.0	5.7
ɛ	8.9	6.4	6.9	8.5	7.2	7.5	6.8	8.9
æ	8.9	12.6	11.1	10.7	12.8	12.2	9.5	11.0
ɝ	3.0	5.7	4.5	5.4	6.5	8.1	5.0	10.0
ɑ	7.8	5.8	5.9	6.0	5.0	6.6	2.7	4.8
ʌ	8.5	10.2	6.5	12.4	8.3	7.7	7.4	8.9
ɔ	3.3	5.4	3.7	3.4	5.5	4.8	4.8	3.4
oʊ	9.3	7.7	11.1	7.5	5.6	7.8	6.5	6.8
ʊ	2.9	1.9	2.3	0.9	2.9	1.8	2.0	0.6
u	8.2	3.3	5.4	3.9	6.6	3.4	9.8	3.6
aɪ	7.5	8.8	8.5	10.4	6.9	7.2	8.8	8.8
aʊ	2.5	2.5	2.9	2.4	3.7	2.8	4.7	3.6
ɔɪ	0.7	0.7	1.9	0.6	0.3	0.6	0.3	0.8

Table 17: The frequency of vowel categories for each speaker across registers expressed as percentages.

Area of vowel space As in Kuhl et al. (1997) and its replications, this area of the vowel space was defined as that defined by the central tendency of the three point vowels, /i/, /ɑ/, and /u/, in the two-dimensional formant space. This statistic was measured separately for each speaker and each register. These central tendencies were defined as the mean value for F_1 and F_2 in Hz of each category. In order to allow for greater comparison with other studies, this statistic considered raw formant values and these areas were reported in Hz^2. The observation of a larger vowel space area in IDS relative to ADS would provide evidence of contrast enhancement in this register. The degree of expansion across registers was also expressed as the ratio of areas across registers with values greater than 1 indicating the presence of hyperarticulation in IDS.

Central tendencies Measures of the mean value of F_1 and F_2 contributed to comparisons of discriminability and maternal affect across registers. Measures of the central tendency of individual categories indicated register-specific differences in discriminability through measures of the peripherality of each vowel category. This measure determined whether the effects that were indicated by measures of the area of the vowel space could be generalised from the point vowels to all vowels in the system. Greater peripherality in IDS has also been associated with greater precision in caregivers' speech

(Bernstein Ratner, 1984). This statistic was defined as the Euclidean distance between the central tendency of a given category and the centre of the vowel space. These mean values were calculated using formant values that were z-scored for each individual speaker: all of the acoustic data that is presented in this chapter was scaled in this way, except for measures of the area of the vowel space. The centre of the vowel space was defined as the grand mean of the fifteen vowel categories. The observation of greater peripherality in IDS relative to ADS would provide evidence of contrast enhancement in this register.

Differences in maternal affect across registers were operationalised by considering whether there was a global difference in central tendency of F_1 and F_2 for all fifteen categories across registers. A global increase in the frequency of either formant in IDS relative to ADS would provide evidence of greater positive affect in this register rather than a pedagogical effect (Benders, 2013).

Dispersion The dispersion of categories in acoustic space was operationalised as the Euclidean distance between the central tendencies of paired vowels. This statistic has been adopted as a measure of discriminability since it can be directly related to the quality of distributional information in the acoustic input (Bohn, 2013; Cristia and Seidl, 2014). As with measures of peripherality, inter-category Euclidean distances were calculated using z-scored category means for F_1 and F_2. Greater dispersion in IDS relative to ADS would provide evidence of contrast enhancement in this register.

Within-category variance Measures of the standard deviation of F_1 and F_2 for each category described differences in the limits of the distributions associated with vowel categories across register. This statistic captured differences in the variability of caregivers' speech across registers. It was reported in order to replicate the common observation that IDS has greater within-category variance than ADS across a range of languages (American English, Russian and Swedish: Kuhl et al., 1997; American English: Cristia and Seidl, 2014, Kirchhoff and Schimmel, 2005, McMurray et al., 2013; Dutch: Benders, 2013; Japanese: Miyazawa et al., 2017). Again, the standard deviation was calculated using z-scored values for F_1 and F_2. The observation of greater within-category variance in IDS relative to ADS would provide evidence of contrast deterioration, rather than enhancement, in this register. All else being equal, greater variance results in greater category overlap and poorer distributional information. This statistic also aimed to facilitate comparisons between measures of dispersion and the degree of overlap.

Degree of overlap The degree of overlap between categories was first observed by plotting a series of ellipses for each category that represent 80% confidence regions in two-dimensional formant space. This method served as a visual indicator of the degree of overlap between categories in each register. Register-specific differences in the degree of overlap between categories were formally operationalised using measures of D(a) (Newman, Clouse, and Burnham, 2001). This statistic indicates a measure of Euclidean distance between the central tendencies of two vowel categories which is sensitive to the within-category variance of each of those vowels. The one-dimensional version of this statistical measure is defined below, where A and B are two vowel categories which are modeled as normal distributions with a mean, μ, and a standard deviation, σ:

$$D(a) = \frac{\sqrt{2}(\mu_A - \mu_B)}{{\sigma_A}^2 + {\sigma_B}^2}$$

This chapter reported the degree of overlap between multidimensional distributions: specifically, vowel categories were considered as two-dimensional distributions of F_1 and F_2. To calculate the degree of overlap in this space, I follow the method used in (Cristia and Seidl, 2014) where multidimensional D(a) was defined as the root sum square of the values of D(a) for each of the individual dimensions. Measures of D(a) were calculated using z-scored values of F_1 and F_1 and this statistic was reported for both of these formants individually, as well as for the two-dimensional space formed by combining them. In comparison to measures of dispersion, this measure provided more transparent evidence of the quality of the distributional information in the input. The value of D(a) is maximal for categories that are distal in acoustic space and that have low within-category variance. Because of this, the observation of greater values of D(a) in IDS relative to ADS would provide evidence of contrast deterioration, rather than enhancement, in this register.

Differences in orientation Differences in the shape and orientation of distributions were measured using the metric S2 (Garcia, 2012). S2 depends on eigenvectors, a representation of orientation that indicates the direction along which a category has its maximum variance. The S2 of two categories A and B can be calculated using the equation indicated below. In this equation, V_{AA} indicates the variance of A along its own eigenvector (that is to say, its eigenvalue), V_{BA} refers to the variance of B which is explained by the eigenvector of A, and so on:

$$S2 = [({V_{AA}}^2 + {V_{BB}}^2) - ({V_{AB}}^2 + {V_{BA}}^2)]^2$$

Greater values of S2 indicate that categories have large differences in orientation relative one another. Therefore, the observation of a greater value for S2 in IDS relative to ADS would provide evidence of effects of contrast enhancement which have been proposed as a potential feature of IDS (Eaves Jr. et al., 2016). However, this type of effect must be interpreted with caution: greater within-category variance necessarily leads to larger eigenvectors and thus larger values of S2. In order to control for this, measurements of S2 were compared with the square of the difference between the first and second eigenvalues of each category. Since the first and second eigenvectors are orthogonal to one another, this statistic provides the maximum possible value for S2 for a given pair of categories. The quotient of the observed value of S2 and this maximal value indicates the extent to which categories are oriented orthogonally to one another: values close to 0 indicate that categories have the same orientation while values close to 1 indicate orthogonal orientations. Thus, greater values for this ratio in IDS relative to ADS would provide evidence of an effect of enhancement in this register.

3.2 Results

3.2.1 Area of the vowel space

Measures of the area of the vowel space are presented in table 18 for each speaker across register. The ratios of ADS to IDS which are presented in this table indicated that each speaker exhibited an effect of vowel space expansion in IDS.

speaker	IDS Area	ADS Area	ratio
ALI	181.1	119.3	1.518
ANN	99.6	92.7	1.062
CIN	252.6	121.1	2.410
GAI	182.7	194.9	1.564

Table 18: The area of vowel spaces in kHz^2 in ADS and IDS, as defined by the mean values for F_1 and F_2 of the three point vowels, /i/, /ɑ/ and /u/. Ratios of the area in IDS to the area in ADS provided evidence of vowel space expansion in IDS for all four speakers.

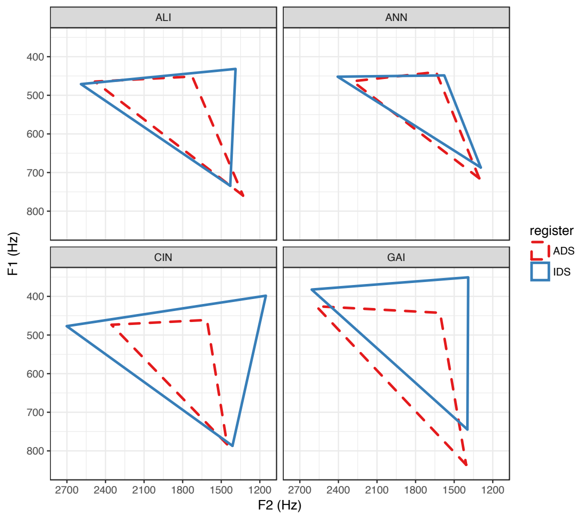

Figure 5: Illustrations of the area of the vowel space in IDS and ADS for each of the four speakers, indicating the effect of expansion in IDS.

Figure 5 further illustrates how the area of the vowel space differed across registers and indicates how the quality of each point vowel affected the value of this composite measure. The effect of expansion in IDS that was seen for speakers ALI and CIN was

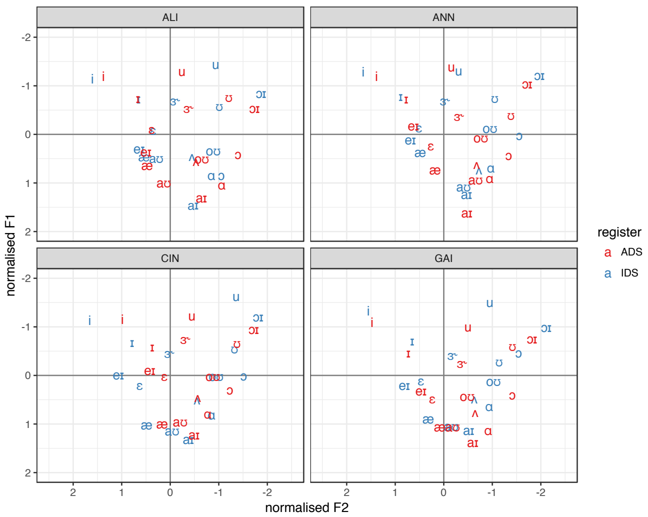

Figure 6: Mean values for F_1 and F_2 for each of the fifteen vowels across speakers and registers.

the result of a greater distance between /i/ and /u/ in the F_2 dimension in IDS relative to ADS. The central tendency of /ɑ/ was comparable across registers for these two speakers. Speaker GAI also showed evidence of expansion in IDS and this results could again be associated with a greater distance between /i/ and /u/ in IDS relative to ADS. Additionally, this speaker showed a global decrease in the value of F_1 in IDS, indicating that point vowels were realised with a closer quality in this register. The smaller effect of expansion which was observed for speaker ANN followed from the fact that this speaker did not make large modifications to the central tendency of any of the point vowels in IDS.

3.2.2 Central tendencies & peripherality

Register-specific differences in the central tendency of vowel categories were inspected visually as demonstrated in in figure 6. For speaker ALI, the peripheralisation of /i/ and /u/ which resulted in an expanded vowel space in IDS was not observed for all of the vowels in the system. For example, /ɪ/ and /ɛ/ had a similar quality across registers and /ɑ/, /ʊ/, and /ɝ/ had a more central quality in IDS. For speaker ANN, IDS vowels were realised with a closer quality than their ADS counterparts, as indicated by the lower values that were observed for F_1 in this register. Though greater peripherality was observed for /i/, /eɪ/, /æ/ and /ɔ/ in IDS compared to ADS, /ʊ/ and /ɑ/ had a more central quality in IDS than in ADS. A set of three front vowels, /ɛ/, /eɪ/, and /æ/, also showed lesser dispersion in IDS than ADS. Speaker CIN showed the most consistent evidence of greater peripherality in IDS than ADS. Each of the five front vowels were had a more advanced quality in IDS than ADS while the back vowels /u/ and /ɔ/ were more retracted in this register. Despite this, the realisation of /ɑ/, /ʌ/, and /oʊ/ was comparable across registers. Speaker GAI tended to realise vowels with

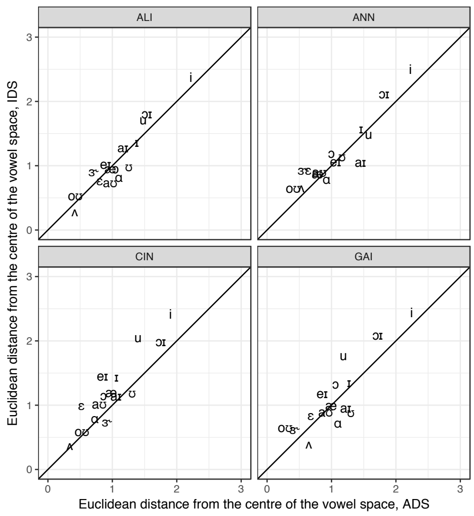

Figure 7: Comparisons of the Euclidean distance of each vowel from the centre of the vowel space in z-scored formant space across speakers and registers. Since measures from IDS and ADS are plotted against each other, data points above the line indicate that a specific category was more peripheral in IDS than ADS. Greater peripherality was observed in IDS for speaker CIN.

a closer quality in IDS than ADS, as indicated by lower values for F_1 in this register. This speaker showed some evidence of peripheralisation as /eɪ/ /ɛ/, and /æ/ were more peripheral in IDS than ADS while /u/ and /oʊ/ were more retracted in this register. Greater peripherality was not observed globally in IDS since /aɪ/, /ɑ/, and /ʊ/ had a more central quality in this register relative to ADS.

In addition to these descriptive generalisations, the peripherality of the fifteen vowels of American English was objectively compared across registers with a series of Wilcoxon signed-rank tests. Figure 7 indicates register-specific differences in the Euclidean distance between the central tendency of each vowel category and the central of the vowel space. For three of the four speakers, the degree of peripherality did not differ across the two registers (ALI, W = 57, p = .890, 95% CIs [-0.082, 0.125]; ANN, W = 28, p = .083, 95% CIs [-0.007, 0.211]; GAI, W = 43, p = .359, 95% CIs [-0.088, 0.270]). The degree of peripherality was greater in IDS for speaker CIN (W = 11, p = .003, 95% CIs [0.078, 0.377]).

3.2.3 Global differences

A series of Wilcoxon signed-rank tests tested whether there were global differences in the mean value of each of the first two formants across the registers. Register-specific differences in the mean value of F_1 are indicated in figure 8. These statistical tests indicated that vowel categories had a lower mean value of F_1 lower across the board in IDS for speakers ANN (W = 19, p = .018, 95% CIs = [-0.271, -0.023]) and GAI (W = 11, p = .003, 95% CIs = [-0.372, -0.153]), providing evidence against formant raising in IDS relative to ADS. This result indicated that vowels were more close in IDS.

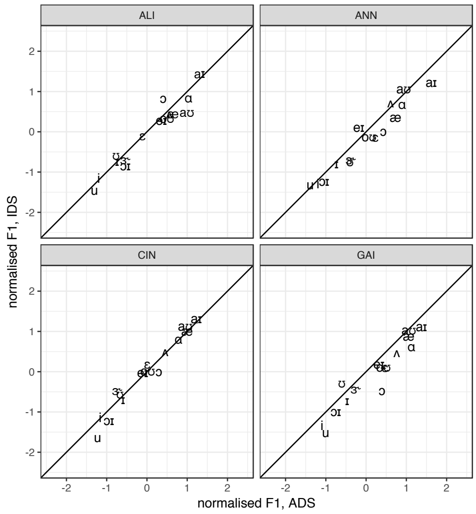

Figure 8: Comparisons of the mean value of F_1 for each vowel category across speakers and registers. Data points above the line have a more open quality in IDS while those below are more close. None of the speakers showed evidence of formant raising in IDS which would be consistent with greater positive affect.

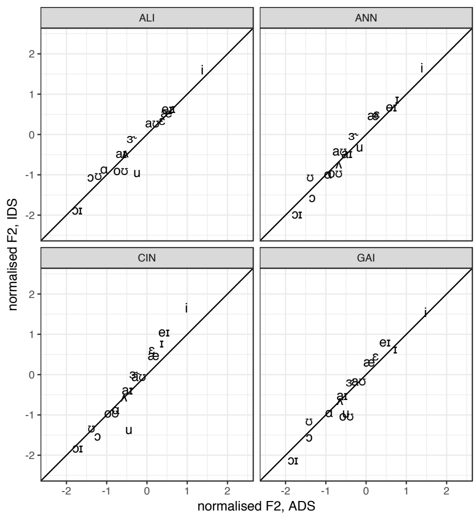

Figure 9: Comparisons of the mean value of F_2 for each vowel category across speakers and registers. Data points above the line have a more advanced quality in IDS while those below are more retracted. Again, none of the speakers showed evidence of formant raising in IDS.

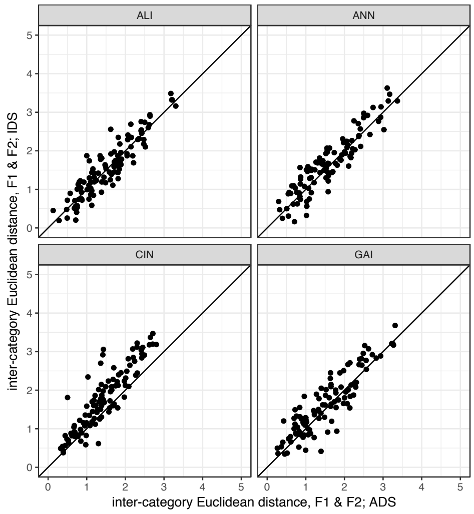

Figure 10: Comparisons of inter-category Euclidean distance in two-dimensional formant space for each of the 105 vowel distinctions across speakers and registers. Data points above the line indicate greater dispersion in IDS. This facilitative effect was observed in IDS across all four speakers.

The mean value of F_1 did not differ across registers for speakers ALI ($W = 38$, $p = .229$, 95% CIs = [-0.174, 0.044]) and CIN ($W = 48$, $p = .524$, 95% CIs = [-0.116, 0.106]).

A parallel set of statistical tests were applied to the measures of the mean value of F_2 in each register which are presented in figure 9. Wilcoxon signed-rank tests indicated that that there was no global difference in the realisation of this formant across registers for any of the four speakers (ALI, $W = 35$, $p = .168$, 95% CIs = [-0.036, 0.192]; ANN, $W = 36$, $p = .188$, 95% CIs = [-0.057, 0.191]; CIN, $W = 35$, $p = .168$, 95% CIs = [-0.078, 0.337]; GAI, $W = 55$, $p = .804$, 95% CIs = [-0.150, 0.153]).

3.2.4 Dispersion

Register-specific differences in the dispersion of vowel categories were operationalised as the Euclidean distance between the central tendencies of pairs of vowel categories. Comparisons of how this measure differed across registers for the 105 vowel distinctions that were considered in the current analysis are presented in figure 10. Wilcoxon signed-rank tests indicated that IDS vowels showed greater dispersion than those in ADS for speakers ANN ($W = 1862$, $p = .003$, 95% CIs [0.033, 0.155]), CIN ($W = 384$, $p < .001$, 95% CIs [0.263, 0.391]) and GAI ($W = 1673$, $p < .001$, 95% CIs [0.059, 0.196]). Measures of dispersion did not differ across registers for speaker ALI ($W = 2663$, $p = .704$, 95% CIs [-0.048, 0.071]).

Inter-category distances were also considered for each of the first two formants individually. Comparisons of the dispersion for the first formant across registers are displayed in figure 11. F_1 only found greater separation in IDS for speaker CIN ($W = 1672$, $p < .001$, 95% CIs [0.039, 0.139]). The three other speakers showed a comparable degree of dispersion for this dimension across registers (ALI, $W = 2657$, $p = .690$,

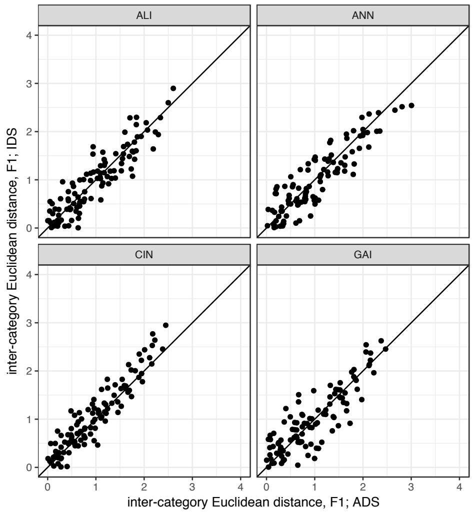

Figure 11: Comparisons of inter-category Euclidean distance for F_1 for each vowel distinction across speakers and registers. Data points above the line indicate greater dispersion in IDS.

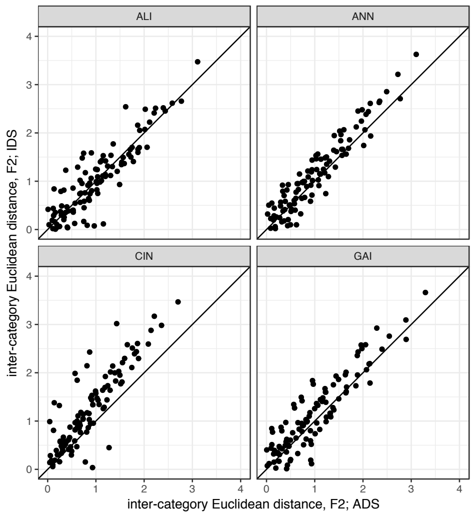

Figure 12: Comparisons of inter-category Euclidean distance for F_2 for each vowel distinction across speakers and registers. Data points above the line indicate greater dispersion in IDS.

95% CIs [-0.065, 0.048]; ANN, $W = 2340,\ p = .158$, 95% CIs [-0.107, 0.0194]; GAI, $W = 2545,\ p = .449$, 95% CIs [-0.035, 0.084]).

Measures of dispersion for the second formant are compared across registers in figure 12. These comparisons indicated that IDS had greater inter-category distances than ADS for three of the four speakers (ANN, $W = 1291,\ p < .001$, 95% CIs [0.086, 0.190]; CIN, $W = 380,\ p < .001$, 95% CIs [0.282, 0.416]; GAI, $W = 1499,\ p < .001$, 95% CIs [0.079, 0.207]). As with other measures of dispersion, inter-category distances for this dimension did not differ across registers for speaker ALI ($W = 2657,\ p = .690$, 95% CIs [-0.046, 0.074]).

3.3 Within-category variance

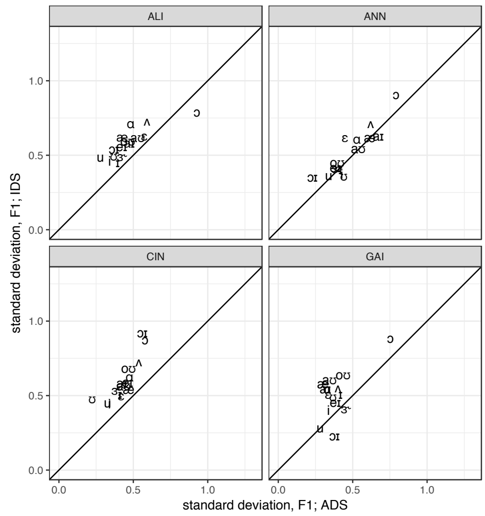

Figure 13: Comparisons of the standard deviation of F_1 for each vowel category across speakers and registers. Data points above the line indicate greater within-category variance in IDS. The IDS productions of speakers ALI, CIN, and GAI were more variable than their ADS productions.

Differences in the variability with which speakers realised vowels across IDS and ADS were operationalised through measures of the standard deviation of each of the first two formants. Register-specific differences in the standard deviation of the first formant are displayed in figure 13 and were tested using Wilcoxon signed-rank tests. The standard deviation of F_1 was greater in IDS vowel production for three of the four speakers (ALI, $W = 9,\ p = .002$, 95% CIs [0.086, 0.164]; CIN, $W = 0,\ p < .001$, 95% CIs [0.118, 0.218]; GAI, $W = 10,\ p = .003$, 95% CIs [0.059, 0.190]). The standard deviation of F_1 did not differ across registers for speaker ANN ($W = 28,\ p = .073$, 95% CIs [-0.005, 0.076]).

Differences in the standard deviation of the second formant across registers are illustrated in figure 14. The standard deviation of this acoustic dimension was greater in IDS than in ADS for the same set of three speakers (ALI, $W = 25,\ p = .048$, 95% CIs

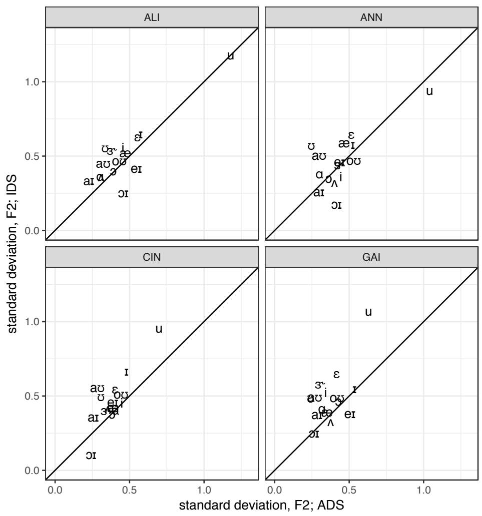

Figure 14: Comparisons of the standard deviation of F_2 for each vowel category across speakers and registers. Data points above the line indicate greater within-category variance in IDS. Again, the IDS productions of speakers ALI, CIN, and GAI were more variable than their ADS productions.

[0.004, 0.107]; CIN, $W = 10,\ p = .003$, 95% CIs [0.032, 0.147]; GAI, $W = 15,\ p = .008$, 95% CIs [0.031, 0.209]). The standard deviation of F_2 did not differ across registers for speaker ANN ($W = 48,\ p = .525$, 95% CIs [-0.050, 0.095]).

3.4 Degree of overlap

Measures of category dispersion and within-category variance across registers provided conflicting evidence with the former indicating enhancement and the latter indicating deterioration. Differences in degree of overlap were therefore considered across registers to determine the relative strength of these two effects. Figure 15 indicates that the vowel categories showed a considerable degree of overlap regardless of the speaker or registers that they were sampled from. This suggested that native language distinctions could not be trivially identified through the use of distributional learning in infancy.

The degree of overlap between vowel categories was operationalised using measures of D(a) in two-dimensional formant space. Comparisons of this variance-sensitive measure of discriminability are illustrated in figure 16. Wilcoxon signed-rank tests indicated that the value of D(a) was significantly lower in IDS in comparison to ADS for speakers ALI ($W = 1125,\ p < .001$, 95% CIs [-0.727, -0.363]) and GAI ($W = 1158,\ p < .001$, 95% CIs [-0.872, -0.410]), indicating greater overlap in this register and thus evidencing an effect of deterioration. The degree of overlap did not differ across registers for speakers ANN ($W = 2613,\ p = .589$, 95% CIs [-0.130, 0.218]) or CIN ($W = 2629,\ p = .625$, 95% CIs [-0.179, 0.126]).

As with the measures of inter-category Euclidean distance, register-specific differences in D(a) were considered for each of the first two formants individually. Figure 17

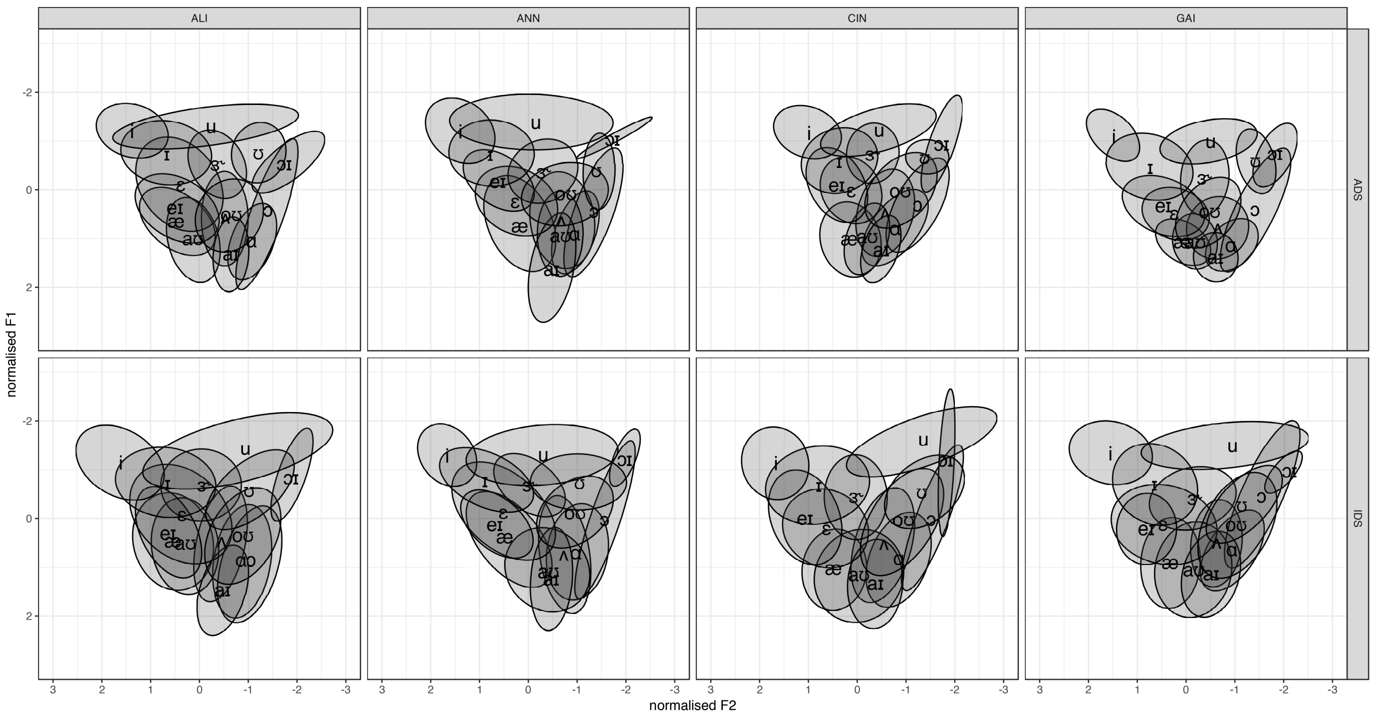

Figure 15: Formant distributions for each vowel in two-dimensional formant space across speakers and registers. The phonetic symbols indicate the central tendency of each category while the ellipses indicate 80% confidence regions for each category. This figure indicates that categories showed a considerable degree of overlap in both IDS and ADS, suggesting that the use of distributional learning may be hindered.

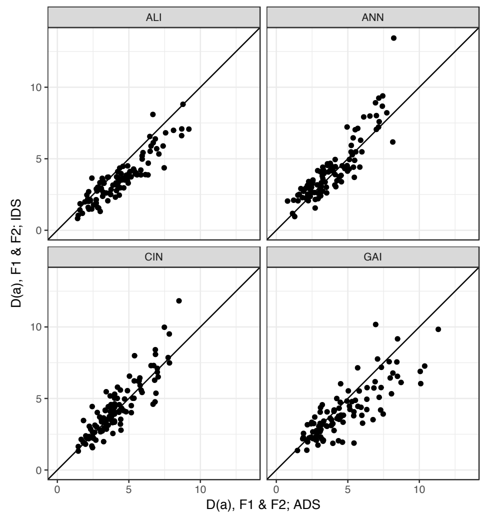

Figure 16: Comparisons of D(a) in two-dimensional formant space for each vowel distinction across speakers and registers. Data points above the line indicate a lesser degree of overlap in IDS. This figure demonstrates a lack of enhancement in IDS relative to ADS: further to this, evidence of deterioration was observed in IDS for ALI and GAI.

indicates how the degree of overlap for F_1 differed across registers. A series of Wilcoxon signed-rank tests indicated the value of D(a) was significantly lower in IDS in comparison to ADS for all four speakers (ALI, W = 986, p < .001, 95% CIs [-0.646, -0.337]; ANN, W = 1770, p = .001, 95% CIs [-0.347, -0.089]; CIN, W = 677, p < .001, 95% CIs [-0.551, -0.326]; GAI, W = 1045, p < .001, 95% CIs [-0.788, -0.398]).

Figure 18 indicates how the degree of overlap for F_2 differed across registers. The value of D(a) was significantly lower in IDS in comparison to ADS for speakers ALI (W = 2067, p = .022, 95% CIs [-0.375, -0.036] and GAI (W = 1997, p = .012, 95% CIs [-0.467, -0.058]). Conversely, the value of D(a) was significantly greater in IDS in comparison to ADS for speakers ANN (W = 1945, p = .008, 95% CIs [0.067, 0.396]) and CIN (W = 1277, p < .001, 95% CIs [0.226, 0.534]).

3.4.1 Relative orientation of categories

Though greater within-category variance results in a greater degree of overlap between categories, one proposal has argued that this property of vowel production may facilitate discrimination if categories are orientated orthogonally to another (Eaves Jr. et al., 2016). Figure 19 indicates how the value of S2 differed across registers: that statistic captures differences in the relative orientation of paired categories. Wilcoxon signed-rank test indicates that the values of S2 was greater in IDS than ADS for speakers CIN (W = 1327, p < .001, 95% CIs [0.012, 0.061]) and GAI (W = 1678, p < .001, 95% CIs [0.013, 0.049]). However, this measure did not differ across registers for speakers ALI (W = 2323, p = .142, 95% CIs [-0.004, 0.046]) and ANN (W = 2239, p = .083, 95% CIs [-0.001, 0.043]).

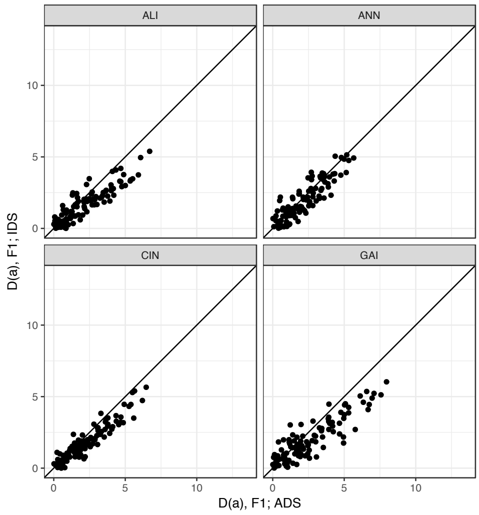

Figure 17: Comparisons of D(a) for F_1 for each vowel distinction across speakers and registers. Data points above the line indicate a lesser degree of overlap in IDS.

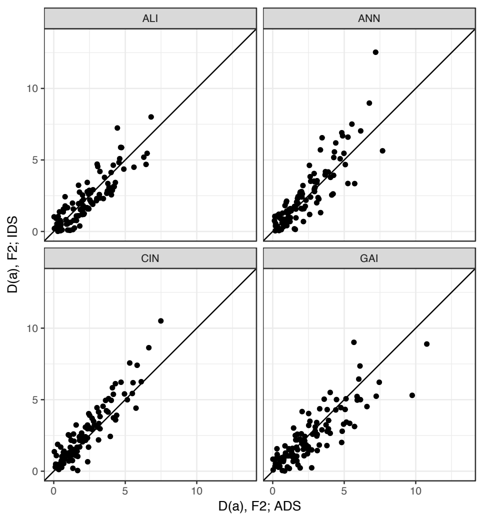

Figure 18: Comparisons of D(a) for F_2 for each vowel distinction across speakers and registers. Data points above the line indicate a lesser degree of overlap in IDS.

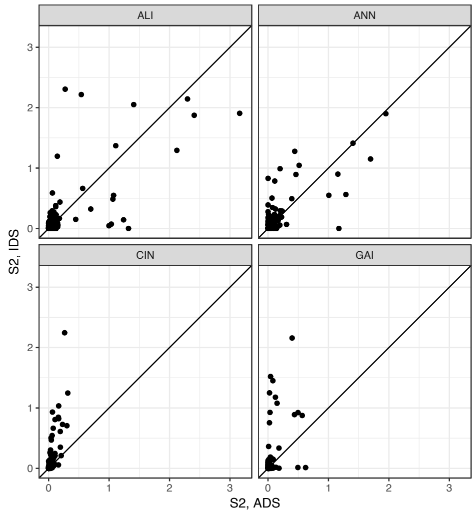

Figure 19: Comparisons of S2 for each vowel distinction across speakers and registers. Data points above the line indicate a greater difference in orientation in IDS. Speakers CIN and GAI showed greater values for S2 in IDS than ADS.

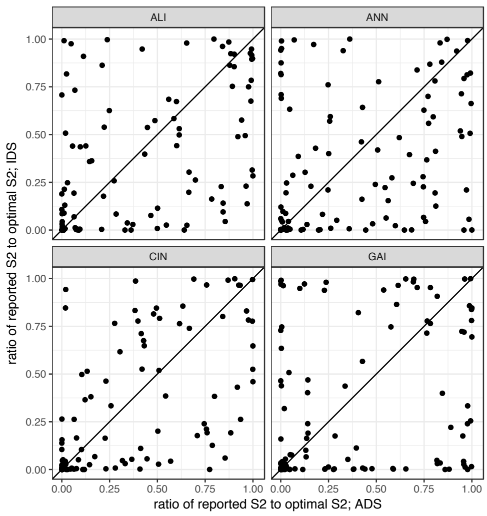

Figure 20: Comparisons of the ratio of observed to maximal S2 for each vowel distinction across speakers and registers. Data points above the line indicate a greater difference in orientation in IDS. This ratio did not differ across registers for any of the four speakers, indicating that speakers did not alter the orientation of categories across registers.

Since the value of S2 increases as within-category variance increases, the heightened values of S2 that were observed in IDS could just have reflected the greater variance that was reported for vowels in this register rather than a genuine difference in orientation. Figure 20 indicates how the the ratio of observed S2 to the maximal possible value for S2 differed across registers. This ratio did not differ across registers for any of the four speakers (ALI, $W = 2365$, $p = .183$, 95% CIs [-0.103, 0.014]; ANN, $W = 2316$, $p = .136$, 95% CIs [-0.135, 0.016]; CIN, $W = 2580$, $p = .518$, 95% CIs [-0.075, 0.022]; GAI, $W = 2447$, $p = .284$, 95% CIs [-0.130, 0.021]).

3.5 Discussion

The current comparative analysis of vowel quality in IDS and ADS considered the productions of four speakers which were sampled from a large, naturalistic speech corpus. This analysis considered over a thousand vowel tokens from each speaker's productions of each register. This analysis operationalised register-specific differences in discriminability through measures of the central tendency and variance of categories and applied these measures to fifteen American English vowels. A summary of these measures and any register-specific effects that were detected through the use of statistical tests are presented in table 19. Measures of the area of the vowel space and dispersion indicated that the adjustments that speakers made to the central tendency of vowels in IDS were consistent with an effect of enhancement. However, variance-sensitive measures did not support the hyperarticulation hypothesis: measures of the degree of overlap that indicated the quality of distributional information in the input showed an effect of contrast deterioration in IDS, if anything.

F_1, F_2	ALI	ANN	CIN	GAI
area	I > A	I > A	I > A	I > A
peripherality	ns	ns	I > A	ns
2D dispersion	I > A	I > A	I > A	I > A
F_1 dispersion	ns	ns	I > A	ns
F_2 dispersion	ns	I > A	I > A	I > A
F_1 variance	I > A	ns	I > A	I > A
F_2 variance	I > A	ns	I > A	I > A
D(a)	A > I	ns	ns	A > I
F_1 D(a)	A > I	A > I	A > I	A > I
F_2 D(a)	A > I	I > A	I > A	A > I
S2	ns	ns	I > A	I > A
S2 ratio	ns	ns	ns	ns
mean F_1	ns	A > I	ns	A > I
mean F_2	ns	ns	ns	ns

Table 19: A summary of the acoustic measures of discriminability (and positive affect) for the analysis of F_1 and F_2. This table indicates the presence and direction of any significant effects which were identified through the use of Wilcoxon signed-rank tests.

The effect of vowel space expansion in IDS relative to ADS replicated the original result of Kuhl et al. (1997). Similar facilitative effects have been reported in previous studies of American English IDS (Cristia and Seidl, 2014; Hartman, Bernstein Ratner, and Newman, 2016; Wieland et al., 2015). Comparisons of the peripherality of vowel categories across registers indicated that this effect of expansion could not be generalised to all vowels in the system. Greater peripherality was only observed across the board in IDS relative to ADS for speaker CIN. Each of the other speakers produced vowels with a comparable degree of peripherality across registers. Instead, measures of inter-category Euclidean distances indicated that register-specific differences in the central tendency of categories were consistent with greater dispersion in IDS. All four speakers showed greater dispersion in IDS in comparison to ADS, providing partial support for the hyperarticulation hypothesis. The fact that IDS vowels were not more peripheral than those in ADS aligned with previous studies that have failed to find such an effect of peripheralisation in American English IDS (McMurray et al., 2013). Further to this, the pattern of greater dispersion that was observed in IDS can be contrasted with previous studies which have reported this measure across registers. Previous comparisons of inter-category Euclidean distance have failed to demonstrate an effect of enhancement for a subset of distinctions in samples of IDS from American English (/i, ɪ/, /eɪ, ɛ/: Cristia and Seidl, 2014) and Danish (/i, e/, /eː, ɛː/, /oː, ɔː/: Bohn, 2013).

Though greater dispersion was suggestive of enhancement in IDS, measures of variance must also be considered across registers in order to closely relate these findings to the learning mechanisms that are available in infancy. Distributional learning depends on the existence of a one-to-one relationship between the number of modes in the acoustic input and the number of distinctive categories in a language (Maye, Werker, and Gerken, 2002). As highlighted in Cristia and Seidl (2014), this effect of greater dispersion will only result in a lesser degree of overlap if measures of within-category variance are comparable across registers. The current acoustic analysis indicated that the variance of IDS vowels was greater than ADS vowels in both F_1 and F_2 for speakers ALI, CIN and GAI. Measures of within-category variance did not differ across registers for speaker ANN. These results were consistent with contrast deterioration and conformed with previous acoustic analyses of IDS which have identified high within-category variance as a feature of this register across multiple languages (American English, Russian and Swedish: Kuhl et al., 1997; American English: Cristia and Seidl, 2014, Kirchhoff and Schimmel, 2005, McMurray et al., 2013; Dutch: Benders, 2013; Japanese: Miyazawa et al., 2017).

Measures of D(a) were reported to resolve these conflicting effects of dispersion and variance. IDS vowels were found to have a greater degree of overlap than their ADS counterparts for speakers ALI and GAI. The observation that measures of D(a) did not differ across registers for ANN and CIN also indicated a lack of enhancement in IDS, in spite of the fact that these speakers showed evidence of greater dispersion in this register. These variance-sensitive measures aligned with previous studies that have applied measures of overlap to samples of IDS and ADS. An analysis of two

American English tense-lax distinctions found that the degree of overlap for /eɪ, ɛ/ was comparable across registers while /i, ɪ/ had a greater degree of overlap in IDS and ADS (Cristia and Seidl, 2014). A comparable analysis of the ten distinctions between the five vowels of Japanese found that these categories had a comparable degree of overlap across registers (/i/, /ɛ/, /a/, /o/, and /ɯ/: Miyazawa et al., 2017). Both of these registers had a greater degree of overlap than clear speech. The current analysis extended these analyses by providing evidence of a lack of enhancement in IDS across a large number of contrasts that were sampled from naturalistic input. As well as allowing for this lack of enhancement to be stated with greater generality, the current analysis suggested the absence of facilitative effects in IDS was not affected by the relative positions of the selected categories in acoustic space or by the type of featural contrast that they represent.

This analysis further explored the possibility that speakers modified distinctions in height and backness in different ways in IDS vowel production. Only one speaker, CIN, showed greater dispersion in F_1 in IDS than ADS and all speakers had a greater degree of overlap in IDS for this acoustic dimension, By contrast, speakers ANN, CIN, and GAI had a greater F_2 dispersion in IDS than ADS while measures of D(a) for F_2 only indicated a similar effect of deterioration in IDS for speakers ALI and GAI. Speakers ANN and CIN showed a lesser degree of overlap for this acoustic dimension in IDS relative to ADS.

Though measures of overlap were not consistent with the facilitation of distributional learning in infancy, the current chapter explored an alternative proposal which states that greater variance may in fact facilitate discrimination in infancy. Heightened variance may allow learners to detect distinctions between high variance vowels if they are oriented orthogonally to one another in acoustic space (Eaves Jr. et al., 2016). Differences in the orientation of categories were operationalised using S2 (Garcia, 2012). Though speakers CIN and GAI did show greater raw values of S2 in IDS than ADS, a further consideration of this statistic did not indicate that there was a genuine difference in the orientation of categories across registers. Because of this, the effects of variance that were observed in this chapter must be interpreted as having a negative effect of the discriminability of categories in IDS.

Since the current analysis considered register-specific differences in vowel quality using data that was originally collected in Bernstein Ratner (1984), it is apt to draw comparisons between the measures of peripherality and overlap that were reported across these two analyses. This is especially important since the chapter stands in opposition to the findings of the original study. Bernstein Ratner (1984) observed that vowels were more peripheral and showed a lesser degree of overlap than in speech to the oldest set of infants in comparison to ADS. A series of methodological differences limit the relevance of the comparisons that can be drawn here. While the current acoustic analysis considered 8,234 IDS vowel tokens and 5,451 ADS vowel tokens from a set of four speakers, the original study considered 2,406 vowel tokens sampled from nine speakers across the two registers. Bernstein Ratner (1984) reported register-specific differences in peripherality and overlap which were aggregated across multiple speakers that were grouped on the basis of the developmental level of their infant addressee. Additionally, measures of peripherality and overlap were captured through a series of descriptive generalisations and these were reported for a set of nine (or fewer) vowels. By contrast, the current study reported objective statistics that considered how each individual speaker realised the relevant set of fifteen vowels.

Though the current set of acoustic results indicated that speakers did not enhance vowel distinctions in IDS, two further factors must be considered in assessments of this register. Firstly, this analysis only considered formant measures of F_1 and F_2 and provided only a limited view of the acoustic dimensions that contribute to vowel identity in American English. One possible interpretation for this lack of enhancement is that formant analyses do not accurately represent the intentions of caregivers in IDS vowel production (Eaves Jr. et al., 2016). If speakers optimise the discriminability of native language distinctions by minimising overlap in high dimensional space, any effects of enhancement may be misinterpreted when viewed in two-dimensional formant space. Under this approach, formant analyses which do not indicate an effect of enhancement in IDS can be interpreted false negatives. The use of multidimensional data has also been advocated as a method of assessing the viability of distributional learning in infancy (Swingley, 2009). Acoustic dimensions beyond the first two formants may help to mitigate the cases of overlap which have been observed in formant analyses of IDS. Secondly, this analysis did not indicate the impact that these register-specific effects may have on the infant learner. Comparisons of the relative discriminability of IDS and ADS are suitable for identifying whether IDS is consistent with enhancement or deterioration. However, the magnitude of these effects remain unclear: deterioration may severely hinder the infant learner or just have a marginal effect on discriminability. Assessing the magnitude of these effects requires the use of computational models which assess the absolute discriminability of vowels in a given sample of input. The impact of any register-specific differences in discriminability can therefore be explored by comparing the performance of these models across registers.

4 Multidimensional acoustic analysis of IDS & ADS

The previous chapter presented a formant analysis which determined the extent to which caregivers alter their realisation of vowels in IDS in order to facilitate perceptual attunement in infancy. Though measures of the central tendency of each category indicated that the vowels of American English IDS were more dispersed in acoustic space than their ADS counterparts, measures of within-category variance were generally greater in the IDS vowel system. This analysis indicated that the degree of overlap between vowels was greater in IDS than ADS, if anything. The distributional properties of IDS therefore did not enable the use of distributional learning in infancy to a greater extent than those of ADS. The modifications that speakers made in this register therefore could not be motivated as having a facilitative effect on the detection and processing of native language vowel distinctions. Moreover, these acoustic results were associated directly with the learning mechanisms that are available in infancy, unlike measures of vowel space expansion.

The analysis in chapter 3 provided new insights into the hyperarticulation hypothesis by applying a diverse set of measures of discriminability to an exhaustive set of categories which were sampled from data which strongly resembles the linguistic input. However, this formant analysis only provided a limited view of the acoustic dimensions that are relevant to vowel quality in American English. Static measures of the first two formants do not adequately describe these distinctions (Hillenbrand et al., 1995; Hillenbrand, 2013). In order to address this, the current chapter applied the same measures of discriminability which were adopted in the previous chapter to measures of the third formant, patterns of spectral change, and vowel duration. Register-specific differences in vowel quality were further explored in the high dimensional acoustic space that is defined by the first two formants in combination with these additional acoustic dimensions.

By presenting a comparative multidimensional acoustic analysis of IDS and ADS vowels, the current chapter extends empirical work in this domain. The consideration of dimensions beyond F_1 and F_2 has been advocated as one way to further current understandings of hyperarticulation in IDS and the use of distributional learning in infancy. One proposal states that multidimensional data may provide a more transparent view of the intentions behind vowel production in IDS (Eaves Jr. et al., 2016). If speakers aim to minimise the degree of overlap between categories in high dimensional space, then previous formant analyses may have provided an incomplete and misleading view of the properties of this register. The consideration of broader set of acoustic dimensions may provide stronger evidence of hyperarticulation. Such a result would indicate that formant analyses which have failed to demonstrate an effect of enhancement are false negatives. Another proposal states that multidimensional data may help to mitigate the cases of category overlap that have been observed in formant analyses of IDS (Swingley, 2009). If acoustic dimensions beyond F_1 and F_2 provide learners with relevant information about category identity, then previous formant analyses may have overstated the ambiguity of the input. Such ambiguous input hinders the use of distributional learning in infancy. Though this claim was originally made in reference to IDS vowel production, this claim is equally applicable to the quality of distributional information in ADS.

4.1 Methodology

The current analysis considered the same set of vowel tokens as in the previous chapter which were sampled from the Bernstein Ratner speech corpus (Bernstein Ratner, 1984). A total of 13,685 vowel tokens were extracted from the speech of four speakers across the two registers. Each of the four speakers were recorded speaking in each register across three recording sessions which were taken at eight-week intervals. A total of 8,234 IDS vowel tokens were uttered a series of unstructured play sessions while a total of 5,451 tokens were extracted from a series of directed interviews that were led by an adult experimenter.

4.1.1 Data extraction and acoustic analyses

As in the previous chapter, the current acoustic analysis considered register-specific differences in the acoustic properties of fifteen American English vowel categories. This set of vowels comprised the twelve vowels which were analysed in Hillenbrand et al. (1995) (/i/, /ɪ/, /eɪ/, /ɛ/, /æ/, /ɝ/, /ʌ/, /ɑ/, /ɔ/, /oʊ/, /ʊ/, /u/) as well as three diphthongs (/aɪ/, /aʊ/, /ɔɪ/). These categories were again uniquely paired to form a set of 105 distinctions. The current analysis considered measures of the third formant, vowel duration and patterns of spectral change. Further to this, vowel quality was also considered in a high-dimensional space which was defined by measures of the first three formants, patterns of spectral change, and vowel duration.

As with the previous formant analyses, the current multidimensional acoustic analysis was partially automated through the use of the FAVE suite (Rosenfelder et al., 2011). Vowel tokens were located in the corpus using forced alignment through `FAVE-align`. Measures of the third formant and the duration of each of these tokens were extracted from the corpus using the default implementation of `FAVE-extract`. Measures of patterns of spectral change, by contrast, were extracted using a set of scripts which implemented an adapted version of the `FAVE-extract` methodology. I generated a series of candidate analyses using Praat scripts (Boersma and Weenink, 2018) and selected optimal candidates from amongst these using a process of Bayesian inference which I implemented in `R` (R Core Team, 2013). The current section will describe the use of the FAVE suite and motivate the use of a specialised analysis for patterns of spectral change. Moreover, this section will describe how the measures of discriminability from the previous chapter were applied to these additional acoustic dimensions.

F_3 The third formant is an index of rounding and rhoticity. Back rounded vowels have a low third formant while unrounded vowels have high values. Two American English vowels can be distinguished from the rest of the system through measures of this dimension: the rhotic vowel, /ɝ/, has an extremely low F_3 while the high front unrounded vowel, /i/, has extremely high F_3. I applied `FAVE-extract` to the corpus data in order to measure the value of the third formant of each vowel that that was identified through the use of `FAVE-align`. Since this tool measures F_3 by default, the current analysis of F_3 made use of the methodology that was described in 3.1.2. In short, `FAVE-extract` estimates the third formant by using Bayesian inference to select an optimal LPC analysis from a set of candidate analyses. An optimal analysis is selected by candidate LPC analyses to a prior which consisted of values for the mean and bandwidth of the first two formants. It should be noted that neither the mean nor the bandwidth of F_3 are considered in this process. This method assumes that a set of LPC parameters which successfully identify F_1 and F_2 also provides an appropriate value for F_3. As with other formant measures reported in this thesis, measures of F_3 were z-scored in order to ensure that the range of each acoustic dimension was comparable.

Patterns of spectral change Patterns of spectral change refer to the changes in the value of the first two formants within the duration of a single vowel. These changes define the three diphthongs of American English, /aɪ/, /aʊ/, and /ɔɪ/: the value of F_1 decreases throughout the duration of these raising diphthongs. This property is also relevant for monophthongs as /ɛ/ and /ɔ/ become backer and more open, indicated by an increase in F_1 and a decrease in F_2 (Hillenbrand et al., 1995; Nearey and Assmann, 1986). The current analysis considered measures of the first two formant values which were taken at 20% and 80% of the duration of each vowel token. This choice was motivated by the results of a discriminant analysis of American English vowel production (Hillenbrand et al., 1995). Analyses which included two measures of the first two formants outperformed those that had a single measure. However, diminishing returns were seen when analyses with three measures were compared to those which used two measures.

As in the default implementation of `FAVE-extract`, the adapted methodology which I implemented selected an optimal analysis from multiple candidate analyses through Bayesian inference. Though `FAVE-extract` does measure formants at five different time points by default, I decided to re-implement this procedure in order to address some limitation which I identified in the original method regarding the priors which the suite uses. The default implementation of `FAVE-extract` estimates the value of the first two formants at 20%, 35%, 50%, 65% and 80% of each vowel's duration. At each time point, the first two formants are estimated using the LPC parameters which were selected as optimal for the steady state. This method is limited in that it does not allow for different parameters to be used for each time point. As previously stated, the appropriate set of LPC parameters depends on the identity of the vowel being analysed. Similarly, it may be desirable to use different sets of LPC parameters for the differences in quality which defined patterns of spectral change.

In order to allow for LPC parameters to differ across time points, I implemented an adapted version of the FAVE methodology. This only differed from the default implementation of `FAVE-extract` in how it selected an optimal analysis. Rather than using a single prior which considered the mean and bandwidth of first two formants in the steady state, the current adapted method used priors which consisted of the value and bandwidth of the first two formants at both the 20% and 80% time points. I used Praat scripts to locate the two time points for each vowel token that was identified with `FAVE-extract` and to generate a set of four candidate LPC analyses at each of these points. As in the original method, the four candidate LPC analyses respectively attempted to locate 3, 4, 5, or 6 formants in a range of frequencies between 0 and 5500Hz. Since the Atlas of North American English (Labov, Ash, and Boberg, 2005) does not report typical values for F_1 and F_2 at these time points, it was also necessary to generate a set of prior distributions for each category and speaker. These priors consisted of the distribution of values for the mean and bandwidth of the first two formants across the two time points as estimated by LPC analyses which attempted to locate four formants for the relevant category in the relevant speaker's data. It should be noted that these priors were weak and did not result in analyses with four formants being selected as optimal in the majority of cases. As in `FAVE-extract`, the selection of optimal candidates was an iterative process. In the first run, a single optimal analysis for each token was selected by comparing each of the candidate analyses to its relevant prior using Mahalanobis distances. In each run subsequent run, the optimal candidates from the previous run were used as updated priors. I implemented this process of Bayesian inference such that it iterated until no further changes were made to the set of LPC analyses which were selected as optimal.

For each of the first two formants, patterns of spectral change were defined as the difference in the given formant between the 80% and 20% time points. These two values were then z-scored separately for each of the four speakers. The change in F_1 and F_2 can be considered as individual dimensions or as a two-dimensional space that describes the dynamic movement of formants. Positive values for the change in F_1 therefore indicate that the vowel becomes more open while negative values indicate that the vowel becomes more close. Similarly, positive values for the change in F_2 indicated advancement whereas negative values indicated retraction. A value of zero indicated monophthongal quality in each case.

Vowel duration Vowel duration is associated with tense-lax contrasts in American English with the duration of tense vowels being greater than that of their lax counterparts. Perceptual experiments indicate that listeners are more likely to identify tokens with ambiguous formant value as tense if they have a greater duration (Hillenbrand, 2013).

The FAVE suite was used to automate the measurement of vowel duration for each vowel token that was identified in the corpus. The forced alignment which was carried out through `FAVE-align` located the beginning and end of each vowel token in the corpus. The default implementation of `FAVE-extract` uses the temporal information from the forced alignment to calculate measures of vowel duration. As described in 3.1.2, the boundaries at the beginning and end of each segment were placed by dividing the acoustic signal into a series of 10ms slices. Because of this, values of vowel duration were necessarily rounded to the nearest 10ms in the majority of cases. Segmental boundaries which were corrected by hand stood as exceptions to this limitation. The values for vowel duration that are reported in this chapter were log-transformed and then z-scored.

4.1.2 Measures of discriminability

As with measures of the first two formants, the analysis of F_3, patterns of spectral change and vowel duration considered a range of measures of the relative clarity of acoustic distinctions across IDS and ADS. These measures aimed to capture the distributional properties of vowel categories across registers in order to assess the use of distributional learning in infancy. The central tendency of individual vowel distributions was compared using two statistics. Differences in the patterns of spectral change of individual categories were analysed descriptively across registers while global differences in F_3 and vowel duration were tested statistically. A measure of dispersion captured the distance between paired categories in each of these additional acoustic dimensions. The effects of variance were also observed by reporting the standard deviation of each category and D(a) (Newman, Clouse, and Burnham, 2001) was used to operationalised the degree of overlap between categories.

Central tendencies The previous chapter considered the central tendencies of F_1 and F_2 of individual vowels across registers using two different statistics. The peripherality of vowels in the space was analysed as an index of hyperarticulation while global differences in F_1 and F_2 were analysed as markers of affect. For F_3 and vowel duration, global differences were considered across registers in addition to a descriptive analysis of how caregivers modulates these dimensions across registers. A global increase in F_3 has been associated with heightened positive affect in caregivers' speech (Benders, 2013) while a global increase in vowel duration would align with the slower speech rate observed in IDS (Fernald et al., 1989).

Though changes in F_1 and F_2 can be considered as a two-dimensional space that is parallel to static formant measures, it is not informative to test for global differences in the mean value or peripherality of these two dimensions. Unlike static measures, the periphery of dynamic formant space cannot be obviously associated with effects of hyperarticulation. Though a greater degree of spectral change may result in ‘better’ tokens of diphthongs, the same claim cannot be extended to monophthongs. Global differences in the change in F_1 and F_2 were also not compared across registers since they do not align with an increase in discriminability or positive affect. Since patterns of spectral change could not be described aptly through measures of peripherality or global differences, register-specific differences in patterns of spectral change registers were instead inspected visually and analysed descriptively.

Dispersion Measures of the dispersion of categories in acoustic space were adopted for each of the additional acoustic dimensions. As before, dispersion was operationalised as the Euclidean distance between the central tendencies of paired vowel categories. For F_3 and vowel duration, this metric was calculated using paired z-scored category means. For patterns of spectral change, inter-category distances were calculated both for z-scored means for change in F_1 and F_2 individually and for both of these dimensions considered together. Measures of dispersion were also calculated in the multi-dimensional space that encompassed all of these acoustic dimensions as well as the static measures of F_1 and F_2. Greater inter-category distances in IDS compared to ADS would indicate a pattern of contrast enhancement in this register.

Within-category variance A measure of the standard deviation of each acoustic dimension was considered in order to describe the limits of the distributions associated with vowel categories. These metrics were calculated separately for each dimension using z-scored value. This statistic was primarily reported in order to enable a comparison of measures of dispersion and overlap.

Distributional overlap The degree of overlap between categories was measured using D(a) (Newman, Clouse, and Burnham, 2001), as defined in 3.1.3. As with measures of Euclidean distance, this measure was calculated for F_3, vowel duration, the change in F_1, and the change in F_2. Further to this, multidimensional measures of D(a) were calculated for both dynamic formant measures as well as for the high dimensional space defined by these additional acoustic dimensions and the static formant measures. In these cases, I again followed the method used in Cristia and Seidl (2014) where multi-dimensional D(a) was defined as the root sum square of the values of D(a) for each of the individual dimensions. Contrast enhancement in IDS would be indicated by greater values for D(a) than those reported for ADS.

4.2 Results I: F_3

4.2.1 Central tendencies

Register-specific differences in the central tendency of the distribution of F_3 can be observed in figure 21. These values were inspected visually to in order to observe

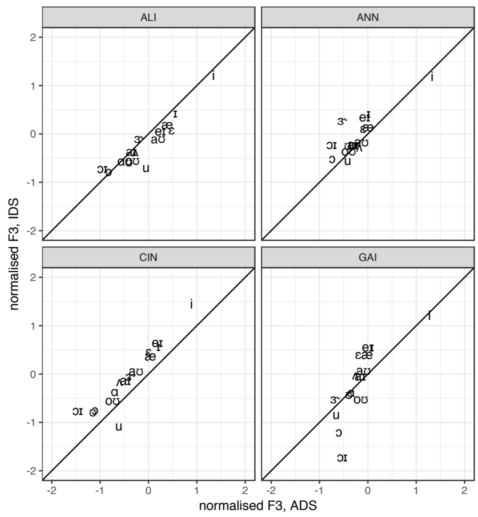

Figure 21: Comparisons of the mean value of F_3 for each vowel category across speakers and registers. Data points above the line have a less rounded quality in IDS while those below are more rounded. Formant raising was observed for ANN and CIN in IDS, consistent with greater positive affect.

patterns of enhancement and differences in affect across registers. Evidence of a lowered third formant for back rounded vowels in IDS would be consistent with enhancement. Comparisons of the central tendency could also provide evidence of formant raising across the board in IDS which has been associated with heightened positive affect in this register (Benders, 2013).

A visual inspection of the central tendencies revealed idiosyncratic patterns in each speaker's production of IDS. Speaker ALI did not provide clear evidence of enhancement. Though /oʊ/, /ʊ/, and /u/ had lower values for F_3 in IDS, the similar cases of lowering that were observed for front vowels /eɪ/, /ɛ/ and /æ/ did not align with contrast enhancement. Speaker ANN had similar values for F_3 for the majority of vowels. Though the raised values for /ɪ/ and /eɪ/ in this speaker's IDS data were consistent with enhancement, similar increases were also observed for /ɔ/, /ɔɪ/ and /ɝ/. Speaker CIN's IDS productions appeared to be consistent with heightened affect in this register. Excluding /u/, every vowel had a greater value for F_3 in IDS than ADS. Speaker GAI's IDS productions showed the strongest evidence of enhancement as the rounded vowels /u/, /oʊ/ and /ɔ/ had a lower F_3 in IDS while the front vowels /eɪ/, /ɛ/ and /æ/ had a higher F_3.

4.2.2 Global differences

The values of central tendency of F_3 which are indicated in figure 21 was compared across registers with a series of Wilcoxon signed-rank tests. The central tendency of F_3 was greater across the board in IDS than ADS for speakers ANN ($W = 17,\ p = .013$, 95% CIs [0.037, 0.326]) and CIN ($W = 11,\ p = .003$, 95% CIs [0.326, 0.460]). This finding was consistent with positive affect in this register. Conversely, the central tendency of F_3 was lower across the board in IDS than ADS for speaker ALI ($W = 23$, $p = .035$, 95% CIs [-0.234, -0.019]). The central tendencies of F_3 did not differ across registers for speaker GAI ($W = 50$, $p = .600$, 95% CIs [-0.215, 0.291]).

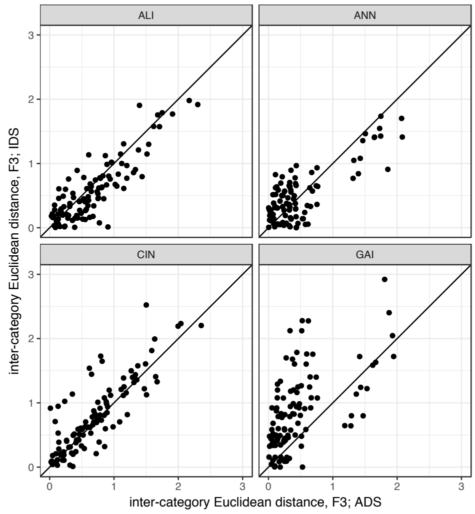

Figure 22: Comparisons of inter-category Euclidean distance for F_3 for each vowel distinction across speakers and registers. Data points above the line indicate greater dispersion in IDS. This facilitative effect was observed in IDS for speakers CIN and GAI while deterioration was seen in speaker ALI's data.

4.2.3 Dispersion

Measures of inter-category Euclidean distances for the third formant are indicated in figure 22. Measures of dispersion for this acoustic dimension were compared across registers through a series of Wilcoxon signed-rank tests. Greater dispersion was observed in IDS than ADS for speakers CIN ($W = 2020$, $p = .015$, 95% CIs [0.010, 0.092] and GAI ($W = 680$, $p < .001$, 95% CIs [0.284, 0.479]). Lesser dispersion was observed in IDS than ADS for speaker ALI ($W = 1637$, $p < .001$, 95% CIs [-0.145, -0.048]). No difference in dispersion was observed for this acoustic dimension across registers for speaker ANN ($W = 2632$, $p = .632$, 95% CIs [-0.045, 0.079].

4.2.4 Within-category variance

Measures of the standard deviation of F_3 are presented in figure 23 and a series of Wilcoxon signed-rank tests were used to compare these values across registers. Within-category variance was lower in IDS than ADS for speaker ALI ($W = 17$, $p = .013$, 95% CIs [-0.180, -0.029]). Within-category variance for this dimension was greater in IDS than ADS for speakers CIN ($W = 21$, $p = .026$, 95% CIs [0.010, 0.176]) and GAI ($W = 3$, $p = .003$, 95% CIs [0.187, 0.401]). No difference in within-category variance was observed across registers for speaker ANN ($W = 34$, $p = .151$, 95% CIs [-0.199, 0.040]).

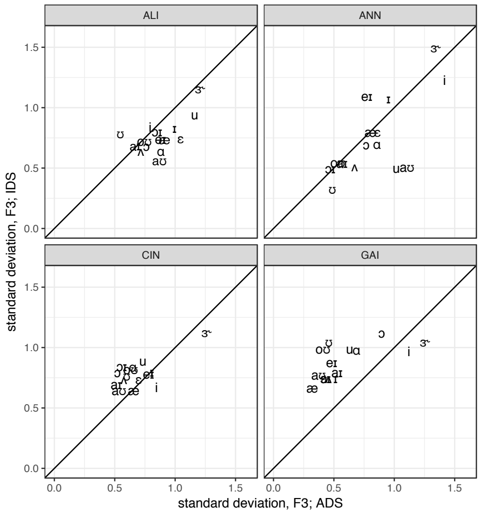

Figure 23: Comparisons of the standard deviation of F_3 for each vowel category across speakers and registers. Data points above the line indicate greater within-category variance in IDS. The IDS productions of speakers CIN and GAI were more variable than their ADS productions.

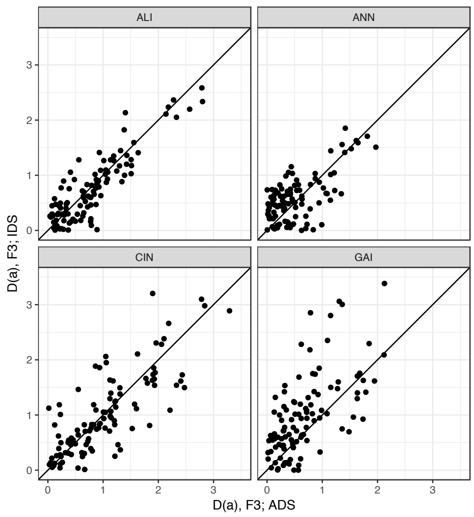

Figure 24: Comparisons of D(a) for F_3 for each vowel distinction across speakers and registers. Data points above the line indicate a lesser degree of overlap in IDS. This figure demonstrates an effect of enhancement in IDS for speakers ANN and GAI.

4.2.5 Degree of overlap

Measures of the degree of overlap were adopted to compare these effects of dispersion and variance and are presented in figure 24. Wilcoxon signed-rank tests indicated that measures of D(a) did not differ across registers for speakers ALI ($W = 2325$, $p = .144$, 95% CIs [-0.097, 0.013]) or CIN ($W = 2533$, $p = .426$, 95% CIs [-0.104, 0.043]). Greater values for D(a), indicating a lesser degree of overlap, were observed in IDS in comparison to ADS for speakers ANN ($W = 2156$, $p = .045$, 95% CIs [0.002, 0.148]) and GAI ($W = 1248$, $p < .001$, 95% CIs [0.164, 0.370]).

4.3 Results II: patterns of spectral change

4.3.1 Central tendencies

Patterns of spectral change were operationalised as the difference in F_1 and F_2 between 20% and 80% of each vowel's duration. The register-specific differences in these acoustic dimensions, as illustrated in figure 25, were first compared through a visual inspection of the central tendency of each category in two-dimensional space. Unlike static measures of F_1 and F_2, the current analysis did not consider global differences in the central tendency or the degree of peripherality of these acoustic dimensions as such patterns do not align with contrast enhancement or heightened affect in IDS.

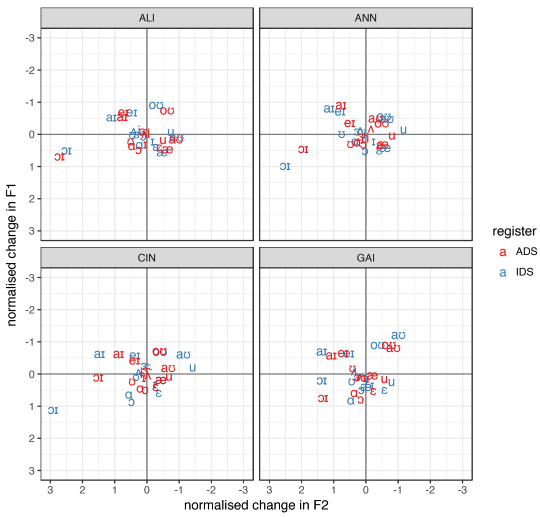

Figure 25: Comparisons of the mean values for the change in F_1 and F_2 for each vowel category across speakers and registers. This plot is arranged such that vowels towards the top left became more close and advanced throughout their duration while those towards the bottom right became more open and retracted.

These measures of spectral change aligned with differences in vowel quality in each register. The vowel /ɔɪ/ became more advanced and open throughout its duration to the extent that formant dynamics distinguished this vowel from all others in the system. Though the phonemic transcription of this vowel indicates that it should become more close throughout its duration, previous descriptions have indicated that the onset of this vowel has a close back quality (Clark and Hillenbrand, 2007). By contrast, the offset was close-mid and advanced in quality. Patterns of spectral change were also relevant for distinguishing /aɪ/ and /aʊ/ from the two low vowels in the system, /ɑ/ and /æ/. Similarly, the advancement and raising of /eɪ/ separated this vowel from other front vowels. The vowel /u/ became backer throughout its duration while /oʊ/ also raised, moving towards a high back quality. These vowels could therefore be distinguished from /ʊ/ and /ɔ/. Lowering was observed for the vowels /ɛ/, /ɔ/, /æ/ and /ɑ/ as indicated by a positive value for the change in F_1. The vowels /i/, /ɪ/, /ɝ/, /ʌ/ and /ʊ/ showed only minimal differences in spectral quality in both IDS and ADS.

Potential patterns of hyperarticulation were identified by visually inspecting the central tendencies of vowels in dynamic formant space in each register. Vowels produced by speaker ALI generally had similar dynamic qualities across registers. However, the vowels /aɪ/ and /oʊ/ showed greater evidence of fronting and raising in IDS. Speaker ANN showed some evidence of enhancement as greater patterns of spectral change were observed in IDS for /oʊ/, /eɪ/ and /ɔɪ/. Additionally, a greater degree of lowering was observed for /ɛ/, /æ/ and /ɔ/ in IDS. The back vowels /u/ and /ʊ/ also respectively showed opposite patterns of backing and fronting in this register. Speaker CIN showed evidence of enhancement in IDS with greater patterns of spectral change in IDS for /eɪ/, /aɪ/, /aʊ/ and /ɔɪ/. Stronger effects of lowered were also observed from /ɛ/, /ɑ/, and /ɔ/ in this register. Similar patterns could be observed for GAI's IDS productions with greater patterns of spectral change for /aɪ/ and /aʊ/. Further to this, patterns of spectral change appeared to be better distinguish /ɛ/ and /ɑ/ from other low vowels in IDS than ADS.

4.3.2 Dispersion

The measures of dispersion for patterns of spectral change in two-dimensional space are presented in figure 26. A series of Wilcoxon signed-rank tests indicated that dispersion was greater in IDS than ADS for ANN ($W = 371$, $p < .001$, 95% CIs [0.246, 0.359]), CIN ($W = 256$, $p < .001$, 95% CIs [0.377, 0.571]) and GAI ($W = 1724$, $p < .001$, 95% CIs [0.049, 0.175]). Measures of dispersion did not differ across registers for speaker ALI ($W = 2365$, $p = .183$, 95% CIs [-0.093, 0.017]).

Figures 27 and 28 indicate measures of dispersion for the change in F_1 and the change in F_2 individually. A pair of parallel analyses demonstrated that register-specific differences in dispersion for the change in F_1 and F_2 individually were consistent with those reported in the two-dimensional analysis. For the change in F_1, a further set of Wilcoxon signed-ranked tests indicated greater dispersion in IDS than ADS for ANN ($W = 1062$, $p < .001$, 95% CIs [0.156, 0.312]), CIN ($W = 1071$, $p < .001$, 95% CIs [0.161, 0.330]), and GAI, $W = 2083$, $p = .025$, 95% CIs [0.009, 0.140]). Speaker ALI showed no difference in dispersion across registers ($W = 2447$, $p = .284$, 95% CIs [-0.026, 0.092]).

For the change in F_2, greater dispersion was observed in IDS relative to ADS for

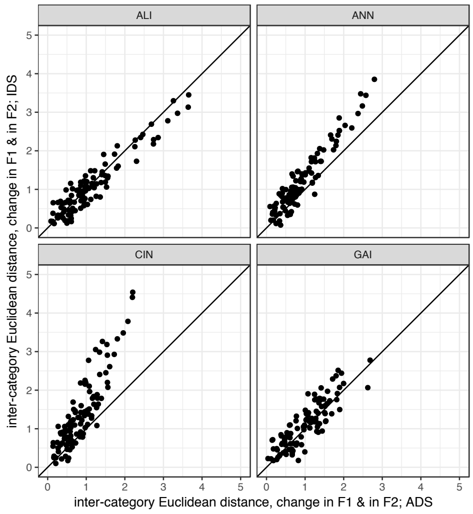

Figure 26: Comparisons of inter-category Euclidean distance for patterns of spectral change for both F_1 and F_2 for each vowel distinction across speakers and registers. Data points above the line indicate greater dispersion in IDS. This facilitative effect was observed in IDS all speakers ANN, CIN, and GAI.

the same three speakers (ANN, $W = 868$, $p < .001$, 95% CIs [0.174, 0.324]; CIN, $W = 990$, $p < .001$, 95% CIs [0.250, 0.484]; GAI, $W = 2042$, $p = .018$, 95% CIs [0.016, 0.161]). Speaker ALI showed no difference in dispersion across registers ($W = 2756$, $p = .934$, 95% CIs [-0.073, 0.065]).

4.3.3 Within-category variance

Wilcoxon signed-rank tests compared the standard deviation of both the change in F_1 and F_2 across registers. Figure 29 indicates the standard deviation of the change in F_1 for each category across registers. Within-category variance was greater in IDS than ADS for ALI ($W = 0$, $p < .001$, 95% CIs [0.181, 0.435]), CIN ($W = 0$, $p < .001$, 95% CIs [0.250, 0.475]) and GAI ($W = 21$, $p = .026$, 95% CIs [0.013, 0.240]). Measures of variance for this acoustic dimension did not differ across registers for ANN ($W = 34$, $p = .151$, 95% CIs [-0.137, 0.020]).

Figure 30 indicates register-specific differences in the standard deviation of the change in F_2. The same set of three speakers showed greater variance for this dimension in IDS (ALI, $W = 4$, $p < .001$, 95% CIs [0.066, 0.240]; CIN, $W = 16$, $p = .010$, 95% CIs [0.059, 0.231]; GAI, $W = 1$, $p = .001$, 95% CIs [0.116, 0.360]) while measures of variance again did not differ across registers for speaker ANN ($W = 53$, $p = .720$, 95% CIs [-0.090, 0.102]).

4.3.4 Degree of overlap

Measures of the degree of overlap were again compared across registers to resolve any conflicting effects of dispersion and overlap. As with Euclidean distances, measures of

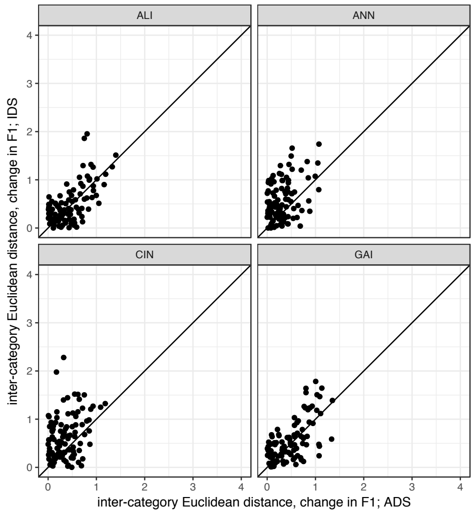

Figure 27: Comparisons of inter-category Euclidean distance for the change in F_1 for each vowel distinction across speakers and registers. Data points above the line indicate greater dispersion in IDS.

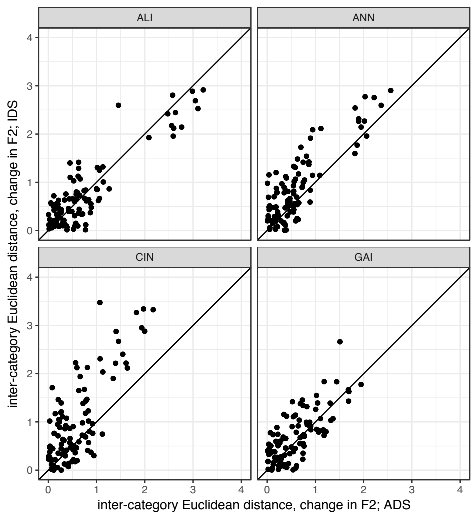

Figure 28: Comparisons of inter-category Euclidean distance for the change in F_2 for each vowel distinction across speakers and registers. Data points above the line indicate greater dispersion in IDS.

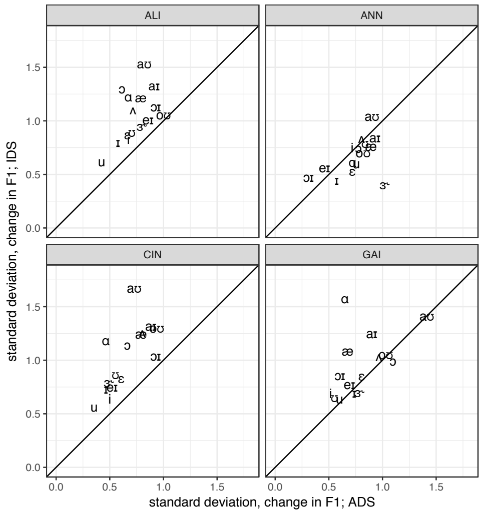

Figure 29: Comparisons of the standard deviation of the change in F_1 for each vowel category across speakers and registers. Data points above the line indicate greater within-category variance in IDS. The IDS productions of speakers ALI, CIN, and GAI were more variable than their ADS productions.

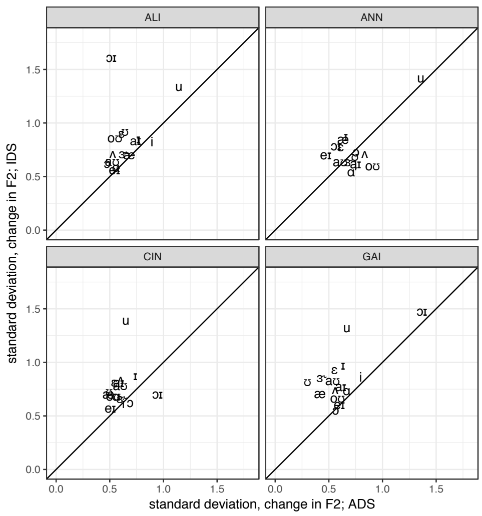

Figure 30: Comparisons of the standard deviation of the change in F_2 for each vowel category across speakers and registers. Again, the IDS productions of speakers ALI, CIN, and GAI were more variable than their ADS productions.

D(a) were first considered for the two-dimensional dynamic formant space and these measures are presented in figure 31. Wilcoxon signed-rank tests indicated that values for D(a) were greater in IDS than ADS for speakers ANN ($W = 309$, $p < .001$, 95% CIs [0.365, 0.503]) and CIN ($W = 1793$, $p = .002$, 95% CIs [0.063, 0.289]). The value of D(a) in IDS was less than that in ADS for ALI ($W = 744$, $p < .001$, 95% CIs [-0.523, -0.273]) and GAI ($W = 969$, $p < .001$, 95% CIs [-0.331, -0.175]).

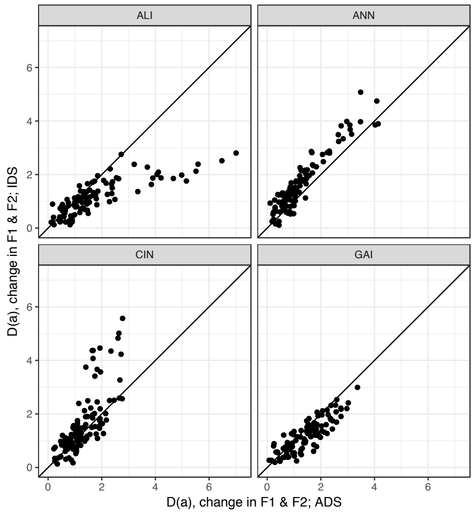

Figure 31: Comparisons of D(a) for patterns of spectral change for each vowel distinction across speakers and registers. Data points above the line indicate a lesser degree of overlap in IDS. This figure demonstrates an effect of enhancement in IDS for speakers ANN and CIN but an effect of deterioration for ALI and GAI.

Further to this, the current analysis considered the degree of overlap for the change in F_1 and the change in F_2 individually, as indicated in figures 32 and 33 respectively. These analyses again revealed results that were similar to the previous two-dimensional analysis. For the change in F_1, the value of D(a) was greater in IDS than ADS for ANN ($W = 720$, $p < .001$, 95% CIs [0.215, 0.360]). The value of D(a) was in lesser in IDS than ADS for ALI ($W = 1276$, $p < .001$, 95% CIs [-0.494, -0.197]) and GAI ($W = 2090$, $p = .027$, 95% CIs [-0.171, -0.009]). By contrast, no difference in overlap was observed across registers for CIN ($W = 2521$, $p = .404$, 95% CIs [-0.102, 0.043]).

For the change in F_2, greater values of D(a) were observed in IDS than ADS for speakers ANN ($W = 720$, $p < .001$, 95% CIs [0.215, 0.360] and CIN ($W = 1376$, $p < .001$, 95% CIs [0.146, 0.387]). Lesser values for D(a) were observed in IDS for ALI ($W = 1276$, $p < .001$, 95% CIs [-0.494, -0.197]) and GAI ($W = 1106$, $p < .001$, 95% CIs [-0.291, -0.144]).

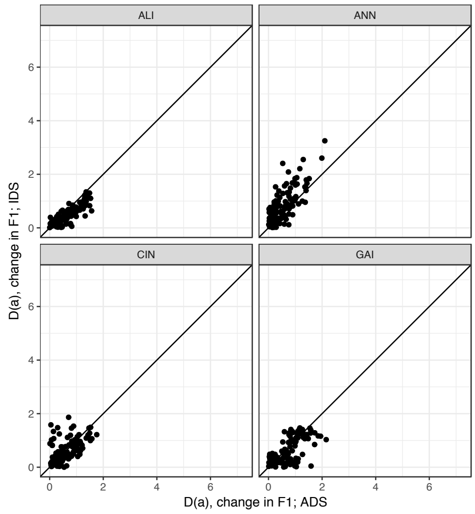

Figure 32: Comparisons of D(a) for the change in F_1 for each vowel distinction across speakers and registers. Data points above the line indicate a lesser degree of overlap in IDS.

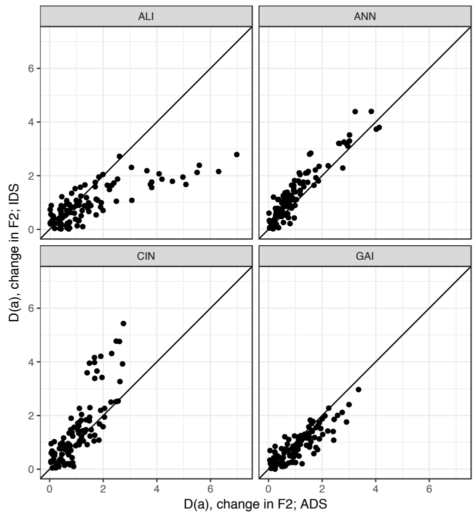

Figure 33: Comparisons of D(a) for the change in F_2 for each vowel distinction across speakers and registers. Data points above the line indicate a lesser degree of overlap in IDS.

4.4 Results III: Vowel duration

4.4.1 Central tendencies

Measures of the log duration were compared across registers in order to observe whether the properties of IDS were consistent with enhancement or a slower speech rate in this register. A visual inspection of the data in figure 34 suggested that IDS had a slower speech rate than ADS, given that the majority of vowels had greater mean values for log duration in this register. In addition to this, the patterns observed across registers for each speaker presented potential evidence contrast enhancement. Though the majority of vowels had a greater duration in IDS than ADS, this effect was more pronounced in tense vowels such as /i/, /oʊ/, /ɔ/ and /ɔɪ/. By contrast, the log duration of lax vowels such as /ɑ/, /ʌ/ and /ʊ/ was either comparable across registers or shorter in IDS than in ADS.

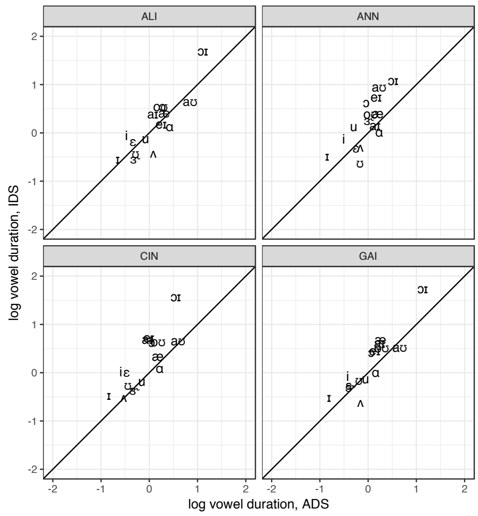

Figure 34: Comparisons of the mean value for log duration of each vowel category across speakers and registers. Data points above the line indicate a slower speech rate in IDS. This effect was observed for speakers ANN, CIN, and GAI.

4.4.2 Global differences

Wilcoxon signed-rank tests determined whether the central tendency of log duration, as illustrated in figure 34, differed across registers. Measures of the central tendency were greater across the board in IDS than ADS for three of the four speakers (ANN, $W = 21$, $p = .026$, 95% CIs [0.060, 0.464]; CIN, $W = 9$, $p = .002$, 95% CIs [0.132, 0.552]; GAI, $W = 27$, $p = .064$, 95% CIs [-0.005, 0.339]). The central tendency of vowel duration showed no global differences across registers for speaker ALI ($W = 47$, $p = .489$, 95% CIs [-0.115, 0.225]).

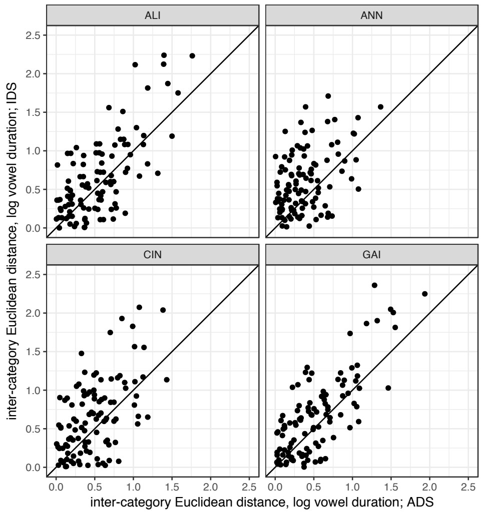

Figure 35: Comparisons of inter-category Euclidean distance for log duration for each vowel distinction across speakers and registers. Data points above the line indicate greater dispersion in IDS. This facilitative effect was observed in IDS across all four speakers.

4.4.3 Dispersion

Figure 35 indicates how measures of the Euclidean distance between the central tendencies of paired categories for log duration differed across registers. Wilcoxon signed-ranked tests indicated that measures of dispersion for this acoustic dimension were greater in IDS than ADS for all four speakers (ALI, $W = 1927$, $p = .006$, 95% CIs [0.029, 0.177]; ANN, $W = 1119$, $p < .001$, 95% CIs [0.142, 0.296]; CIN, $W = 1614$, $p < .001$, 95% CIs [0.077, 0.243]; GAI, $W = 1671$, $p < .001$, 95% CIs [0.060, 0.194]).

4.4.4 Within-category variance

Figure 36 indicates how the standard deviation of log duration differed across registers. Wilcoxon signed-rank tests indicated that within-category variance was greater in IDS than ADS for all four speakers (ALI, $W = 15$, $p = .008$, 95% CIs [0.047, 0.177]; ANN, $W = 21$, $p = .026$, 95% CIs [0.027, 0.211]; CIN, $W = 6$, $p < .001$, 95% CIs [0.077, 0.201]; GAI, $W = 12$, $p = .004$, 95% CIs [0.081, 0.227]).

4.4.5 Degree of overlap

As measures of dispersion and variance presented conflicting effects with regard to the discriminability of distinctions in IDS, this analysis further considered measures of D(a) for this acoustic dimension that are presented in figure 37. Wilcoxon signed-rank tests applied to these measures indicated that there was a lesser degree of overlap in IDS for speaker ANN ($W = 1663$, $p < .001$, 95% CIs [0.075, 0.250]). The degree of overlap in this dimension did not differ across registers for the other three speakers

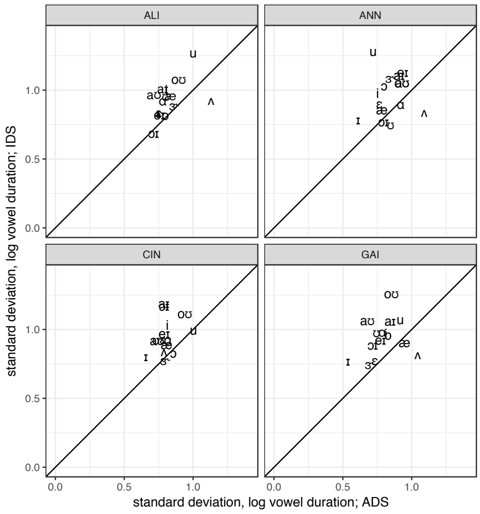

Figure 36: Comparisons of the standard deviation of log duration for each vowel category across speakers and registers. Data points above the line indicate greater within-category variance in IDS. The IDS productions of all four speakers were more variable than their ADS productions.

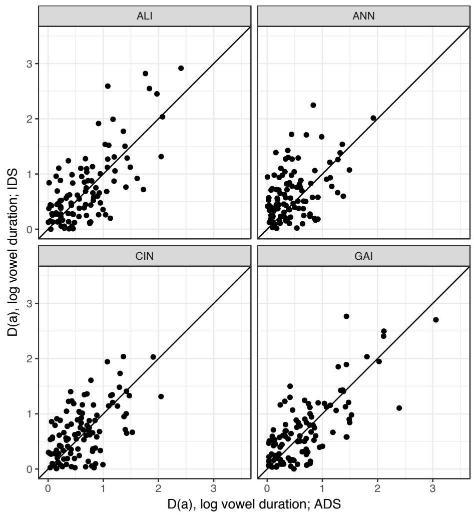

Figure 37: D(a) for log duration for each vowel distinction across speakers and registers. Data points above the line indicate a lesser degree of overlap in IDS. This facilitative effect was observed in IDS for speaker ANN while the other three speakers showed a lack of enhancement.

(ALI, $W = 2494$, $p = .357$, 95% CIs [-0.049, 0.124]; CIN, $W = 2285$, $p = .112$, 95% CIs [-0.017, 0.149]; GAI, $W = 2620$, $p = .605$, 95% CIs [-0.063, 0.104]).

4.5 Results IV: Multidimensional data

The current analysis also considered American English vowel distinctions in the high-dimensional acoustic space that is defined by all of the acoustic measures that have been discussed previously. Specifically, this space consisted of measures of the first three formants, the change in F_1, the change in F_2, and log duration. Considering vowels in high-dimensional space provided insights into how these acoustic dimensions interact. This analysis also addressed the claim that multidimensional acoustic analyses provide the strongest evidence of contrast enhancement in IDS, implying that formant analyses may have mischaracterised the intentions behind caregivers' vowel productions in this register (Eaves Jr. et al., 2016).

4.5.1 Dispersion

Measures of dispersion in high dimensional acoustic space are compared across registers in figure 39. Wilcoxon signed-rank tests indicates that these measures of inter-category Euclidean distances did not differ across registers for speaker ALI ($W = 2692$, $p = .774$, 95% CIs [-0.056, 0.085]). Measures of dispersion were greater in IDS than ADS for the other three speakers (ANN, $W = 702$, $p < .001$, 95% CIs [0.255, 0.415]; CIN, $W = 128$, $p < .001$, 95% CIs [0.455, 0.681]; $GAI, W = 852$, $p < .001$, 95% CIs [0.240, 0.432]).

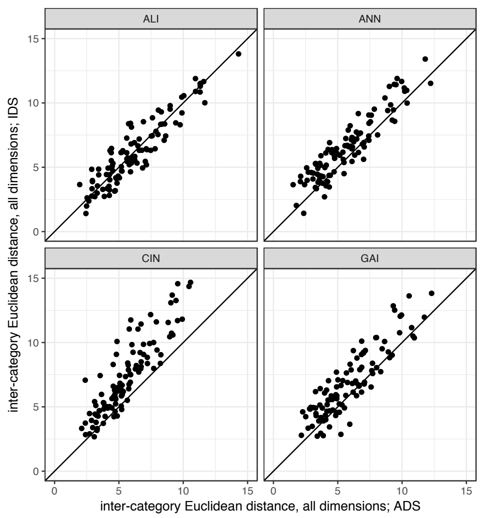

Figure 38: Inter-category Euclidean distances in high dimensional acoustic space for each vowel distinction across speakers and registers. Data points above the line indicate greater dispersion in IDS. This facilitative effect was observed in IDS for speakers ANN, CIN, and GAI.

4.5.2 Degree of overlap

Recall that measures of D(a) for two-dimensional formant space did not indicate that there was a lesser degree of overlap in IDS relative to ADS for any of the four speakers. Figure 38 indicates measures of D(a) in multidimensional acoustic space in order to determine whether additional acoustic dimensions provided stronger evidence of enhancement in this register. Wilcoxon signed-rank tests indicated that the value of D(a) in IDS was less than that in ADS for speakers ALI (W = 451, p < .001, 95% CIs [-0.910, -0.609]) and GAI (W = 1195, p < .001, 95% CIs [-0.813, -0.375]). No difference was observed across registers for speaker CIN (W = 2236, p = .081, 95% CIs [-0.015, 0.310]). Though the results for these three speakers were parallel to those of the formant analysis, stronger evidence of enhancement was observed in multidimensional acoustic space for speaker ANN. Values for D(a) were greater in IDS than ADS for this speaker (W = 1679, p = .004, 95% CIs [0.147, 0.490]).

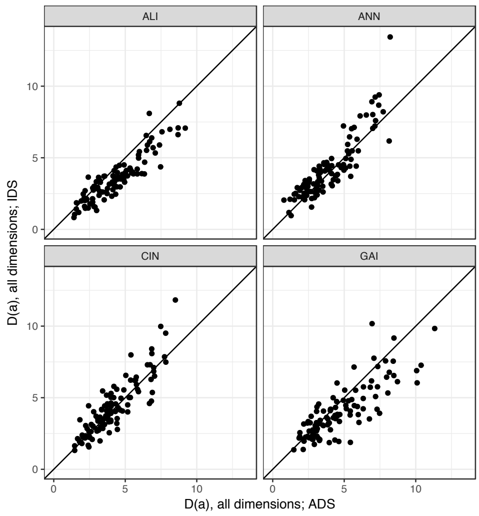

Figure 39: D(a) in high dimensional acoustic space for each vowel distinction across speakers and registers. Data points above the line indicate a lesser degree of overlap in IDS. Though an effect of enhancement was observed in IDS for speaker ANN, deterioration was observed for ALI and GAI.

4.6 Discussion

The current comparative multidimensional acoustic analysis of vowel quality in IDS and ADS extended the formant analysis which was presented in the previous chapter. As before, this analysis considered vowels produced by four caregivers in a naturalistic speech corpus and analysed over a thousand vowel tokens from each speaker's production of each register. Vowel tokens were sampled from each register in order to test the claim that the properties of IDS facilitate perceptual attunement in infancy. As in the previous chapter, this analysis operationalised discriminability through measures of the central tendency and variance of categories and applied these measures

F_3	ALI	ANN	CIN	GAI
dispersion	A > I	ns	I > A	I > A
variance	A > I	ns	I > A	I > A
D(a)	ns	I > A	ns	I > A
mean F_3	ns	I > A	I > A	ns

Table 20: A summary of the acoustic measures of discriminability (and positive affect) that were applied to the third formant. This table indicates the presence and direction of any significant effects which were identified through the use of Wilcoxon signed-rank tests.

spectral change	ALI	ANN	CIN	GAI
dispersion	ns	I > A	I > A	I > A
ΔF_1 dispersion	ns	I > A	I > A	I > A
ΔF_2 dispersion	ns	I > A	I > A	I > A
ΔF_1 variance	I > A	ns	I > A	I > A
ΔF_2 variance	I > A	ns	I > A	I > A
D(a)	A > I	I > A	I > A	A > I
ΔF_1 D(a)	A > I	I > A	ns	A > I
ΔF_2 D(a)	A > I	I > A	I > A	A > I

Table 21: A summary of the measures of discriminability that were applied to patterns of spectral change.

log vowel duration	ALI	ANN	CIN	GAI
dispersion	I > A	I > A	I > A	I > A
variance	I > A	I > A	I > A	I > A
D(a)	ns	I > A	ns	ns
mean log duration	I > A	I > A	I > A	I > A

Table 22: A summary of the measures of discriminability (and speech rate) that were applied to log vowel duration.

all dimensions	ALI	ANN	CIN	GAI
dispersion	ns	I > A	I > A	I > A
D(a)	A > I	I > A	ns	A > I

Table 23: A summary of the measures of discriminability that were applied to the high dimension space that was defined by the first three formants, patterns of spectral change, and log vowel duration.

to fifteen American English vowels. Unlike the previous chapter, these measures of discriminability were applied to measures of the third formant, patterns of spectral change, and vowel duration in addition to measures of F_1 and F_2. Two approaches have advocated for the use of multidimensional acoustic data in analyses IDS vowel production. On the one hand, it has been claimed that multidimensional analyses may provide stronger evidence of contrast enhancement than formant analyses (Eaves Jr. et al., 2016). If speakers optimise the discriminability of native language categories in high dimensional space, analyses which solely consider F_1 and F_2 may have failed to detect these facilitative effects. On the other hand, multidimensional analyses may mitigate the ambiguity of the distributional information that has been observed in formant analyses in IDS and ADS (Swingley, 2009). Such cases of overlap problematically suggest that native language categories cannot be recovered through the use of distributional learning in infancy. If dimensions beyond F_1 and F_2 reliably indicate differences in vowel quality, it is possible that the linguistic input contains statistical regularities which enable the use of distributional learning.

Table 20 summarises the measures of discriminability and the register-specific effects which were identified through the of statistical tests for speakers' realisations of the third formant. Tables 21 and 22 provide similar summaries for patterns of spectral change and log vowel duration while table 23 details analyses in high-dimensional acoustic space. These results broadly aligned with those that were reported in the previous chapter: though IDS vowels showed greater dispersion than their ADS counterparts, within-category variance was also greater in this register. Because of this, comparisons of the degree of overlap across registers generally indicated a lack of enhancement in IDS. Measures of overlap only indicated an effect of enhancement in IDS for speaker ANN when all acoustic dimensions were considered. Speaker CIN show a comparable degree of overlap across registers in this space while an effect of deterioration was observed in the IDS productions of speakers ALI and GAI. The effect of enhancement in multidimensional acoustic space that was observed for ANN was the only case which supported the claim that the consideration of a broader set of acoustic cues provides stronger evidence of hyperarticulation in IDS (Eaves Jr. et al., 2016). The current discussion will now individually address the register-specific differences in the realisation of the third formant, vowel duration and patterns of spectral change which were observed in this chapter.

F_3 As indicated in table 20, register-specific differences in F_3 were first analysed to explore the relative discriminability of vowel production in IDS and ADS. Measures of inter-category Euclidean distance indicated an effect of enhancement in IDS for speakers CIN and GAI. Elsewhere, poorer dispersion was observed for ALI while no difference was observed across registers for ANN. Variance-sensitive measures indicated that IDS vowels had a lesser degree of overlap relative to ADS for ANN and GAI. These effects of enhancement can be contrasted with the measures of overlap that were presented in the previous chapter. None of the measures of D(a) for two-dimensional formant space indicated that vowels in IDS had a facilitative effect on the learner. Measures of D(a) for ALI and CIN aligned with the results of the previous chapter as the degree of overlap did not differ across registers. It should be noted that the lack of enhancement in ALI's IDS data resulted from poorer dispersion and lesser degree of within-category variance in IDS relative to ADS. This pattern is distinct from other cases where greater dispersion and heightened variance resulted in a lack of enhancement in IDS. Further to this, a lesser degree of overlap was observed in IDS for ANN despite the fact that neither dispersion nor variance differed across registers. This effect may therefore have derived from a series of idiosyncratic adjustments to individual categories. This speaker's front vowels had a greater central tendency of F_3 in IDS than ADS while the variance of /ʊ/ and /oʊ/ was lower in this register.

Register-specific differences in the central tendency of F_3 also indicated the affective properties of IDS and ADS. Cases of formant raising were observed for F_3 in the IDS productions of ANN and CIN, indicating greater positive affect in IDS. This was not consistent with the patterns that were observed for F_1 or F_2 for these speakers. The only significant effect from the previous chapter indicated that speaker ANN had lower values for F_1 in IDS. The central tendency of any of the first three formants did was not found to differ across registers for either ALI or GAI. These results were therefore not consistent with the frequency-size relationship (Ohala, 1980, 1984) and did not align with an analysis of Dutch IDS which interpreted the raising of the second and third formants across a set of four vowels as evidence of greater positive affect in this register (/i/, /uː/, /aː/, /ɑ/: Benders, 2013).

Patterns of spectral change Patterns of spectral change were defined as changes in either of the first two formants within the duration of a vowel token. As indicated in table 21, the IDS productions of ANN, CIN and GAI showed greater dispersion relative to ADS for these acoustic dimensions. As with other acoustic dimensions, the within-category variance of these dynamic formant measures was greater in IDS than ADS for speakers ALI, CIN and GAI. Measures of D(a) therefore indicated cases of contrast deterioration in IDS for speakers ALI and GAI. These results strongly resembled the lack of enhancement that was observed in the previous chapter. Unlike the previous analysis, a lesser degree of overlap was observed in IDS for these acoustic dimensions for speakers ANN and CIN. As with the third formant, patterns of spectral change provided limited evidence of enhancement in IDS. This stands in contrast to the formant analyses in which none of the speakers showed an effect of enhancement in this register. Changes in F_1 and in F_2 were also investigated individually: these separated analyses indicated results that were generally comparable to the results of the two-dimensional analysis.

Log vowel duration The analysis of log vowel duration which is summarised in table 22 provided only limited evidence of contrast enhancement in IDS. Speakers ANN, CIN and GAI had greater mean values for this acoustic dimension in IDS than ADS, indicating that speakers had a slower speech rate when addressing infants. Measures of dispersion supported the observation that tense vowels in IDS underwent this effect of lengthening to a greater extent than lax vowels did. Inter-category Euclidean distances were greater in IDS than ADS for each of the four speakers. IDS also had greater within-category variance than ADS across all four speakers. Though speaker ANN had greater values for D(a) in IDS relative to ADS, this measure did not differ across registers for the other three speakers. The observation of a positive effect of dispersion which was negated by increased within-category variance in IDS for this acoustic dimension strongly resembled the results of the formant analyses that were presented in the previous chapter.

General discussion In summary, the current multidimensional acoustic analysis of vowel quality in IDS and ADS provided a set of results which were generally comparable to the formant analysis which was presented in the previous chapter. Though IDS vowel production showed greater dispersion than ADS, greater within-category variance was also observed in this register. The observation of greater variance in IDS for dimensions other than F_1 and F_2 suggested that this is a general property of vowel production in this register. Further to this, measures of the degree of overlap indicated that heightened variance in IDS resulted in a lack of enhancement or even contrast deterioration in IDS. As I have argued previously, this contrast between measures of Euclidean distance and D(a) indicate that assessments of the hyperarticulation hypothesis must adopt measures of discriminability which consider both the central tendency and variance of categories across registers. The current analysis indicated that these patterns occurred across all vowels in the system, suggesting that the discriminability of vowels was independent to the relative position of vowels in acoustic space or the featural distinction that they encoded. Additionally, the current analysis had a high ecological validity as these patterns were observed in a large, naturalistic speech corpus of both IDS and ADS that greatly resembled the input that infants receive on a daily basis.

Though a lack of enhancement was the predominant pattern, this multidimensional analysis did report measures of D(a) which were consistent with an effect of enhancement in IDS. Because of this, the current analysis did provide limited evidence for the claim that the consideration of a broader set of acoustic dimensions may provide strong evidence of enhancement in IDS (Eaves Jr. et al., 2016). For speaker ANN, multidimensional analyses indicated an effect of enhancement that was not observed in two-dimensional formant space. The fact that these two analyses indicated the same results across registers for the other three speakers indicates that formant analyses do provide a reliable indicator of the intentions behind vowel production in IDS. Further to this, the lack of enhancement which was observed in the current analysis aligned with previous multidimensional of these two registers. One analysis of American English operationalised vowel quality using dynamic measures of F_1 and F_2 and vowel duration and considered the realisation of two tense-lax distinctions, /i, ɪ/ and /eɪ, ɛ/, across registers (Cristia and Seidl, 2014). An analysis of Japanese IDS and ADS operationalised acoustic differences through cepstral coefficients which captured the properties of the entire spectral envelop and considered all of the segmental distinctions in this language (Martin et al., 2015). While the American English results demonstrated a lack of enhancement in IDS, the Japanese analysis found a small but significant effect of deterioration across all distinctions. Since multidimensional analyses have not provided stronger evidence of enhancement in IDS, it follows that formant analyses should not be dismissed solely because they do not consider a sufficiently broad set of acoustic dimensions.

The findings of this multidimensional acoustic analysis also have implications for the claim that these additional acoustic dimensions may help to mitigate the cases of overlap which have been observed in formant analyses of IDS (Swingley, 2009). This claim proposes that the absolute discriminability of vowel categories in a sample of data can be increased by considering dimensions beyond the first two formants. Though the original proposal considered vowel production in IDS, increases in discriminability are equally valid for ADS vowel distinctions. This claim makes no specific predictions about the intentions behind vowel production in IDS or the discriminability of this register relative to ADS. Since this claim concerns the absolute discriminability of vowels in multidimensional acoustic space, the comparison of register-specific differences in the current chapter did not directly assess the validity of this claim. The observation of a lack of enhancement in high dimensional IDS data should not be interpreted as evidence that these additional dimensions did have increase the absolute discriminability of distinctions in this register. Instead, the descriptive analysis of the central tendencies of categories and measures of dispersion presented in this chapter provide the most pertinent information regarding this claim. These measures suggest that F_3, vowel duration and patterns of spectral change align with differences in vowel quality in both IDS and ADS and suggest that these dimensions provide learners with additional information that is relevant to the learning task. Comparisons can also be drawn between the measures of dispersion and overlap in two-dimensional formant space and those in high dimensional space. Multidimensional analyses had numerically greater values for Euclidean distance and D(a) than those with fewer dimensions across speakers and registers, suggesting that categories are further separated in higher dimensional space. Care should be taken when interpreting these patterns as these values for these measures will increase by definition as the number of dimensions increases as long as they are applied to two non-identical categories. The computational models in the following chapter will determine whether these additional dimensions mitigate the ambiguity that was apparent in formant distributions. It will present a series of models which resemble the statistical mechanisms that are available to learners in infancy. These models will be applied to acoustic data sampled from both IDS and ADS and the outputs of these models will indicate how these types of input may affect infants' perceptual behaviour within the first year of life.

5 Computational models of perceptual attunement

5.1 Introduction

This thesis addresses two related questions concerning the acoustic properties of vowels in IDS. The first research question considers the extent to which the properties of IDS facilitate the identification of native language distinctions in infancy. The second concerns the viability of distributional learning and the extent to which the observation of statistical regularities in the acoustic input can explain perceptual attunement in infancy. The comparative acoustic analysis of IDS and ADS that was presented in chapters 3 and 4 addressed the first of these research questions. Though this acoustic analysis considered the relative discriminability of vowel categories across registers, it did not describe the absolute discriminability of these categories in each register. The lack of enhancement that was observed in IDS suggested that statistical approaches are not more viable for this register than for ADS. However, the specific impact of these register-specific effects had on distributional learning could not be quantified. The current chapter will therefore present a series of computational models which replicate the statistical learning task. In doing so, it will make explicit predictions about the outcomes of this learning mechanism in infancy and indicate the extent to which vowel categories can be recovered through the use of this mechanism. Clustering techniques replicate the infant learning task as this method assigns individual tokens to an optimal number of groups on the basis of similarity. Logistic regressions assess the predictability of category distinctions on the basis of a series of acoustic dimensions, indicating the reliability of specific distinctions and the relative importance of each dimension. These regressions are a supervised technique and thus should not be viewed as being analogous to the abilities of an infant learner. Instead, these models determine the extent to which an optimal observer would be able to distinguish categories in the input.

By assessing the viability of statistical mechanisms in infancy through the use of a corpus of high dimensional acoustic data that was sampled from both IDS and ADS, this current chapter extended current uses of these modelling techniques. Clustering models which have been applied to entire vowel inventories have shown poor performance (Antetomaso et al., 2016; Feldman et al., 2013; Mooney, 2015). Though this suggests that distributional learning is not a viable mechanism in infancy, models of this type have only been applied to formant distributions which were sampled from either IDS or ADS. By applying these models to multidimensional data from both registers, the current chapter explored whether improved performance was seen in two contexts. Firstly, the current sample of corpus data enabled an evaluation of the claim that multidimensional acoustic data may mitigate the cases of overlap that have been observed in formant distributions sampled from IDS (Swingley, 2009). The acoustic analysis presented in chapter 4 did not indicate whether multidimensional data increased the absolute discriminability of vowels in a single register. Instead, it only demonstrated there was a numerical increase in dispersion and D(a) as further dimensions were included. This claim was assessed by comparing the performance of clustering models applied to low- and high-dimensional acoustic data for each speaker and register. Logistic regressions were also used to explore the role of dimensions beyond F1 and F2 as regression coefficients explicitly indicate how each dimension contributes to the predictability of individual categories. This also represented a novel contribution as previous studies which have applied regressions to samples of IDS and ADS have not used this method to determine how multiple acoustic dimensions predict vowel quality across an inventory as a whole (Bion et al., 2013; McMurray et al., 2013; Werker et al., 2007). By not considering the entire inventory, these regressions may have overstated the discriminability of individual categories since many confusable alternatives were excluded from the analysis.

Secondly, this analysis also considered register-specific differences in model performance, despite the fact that the acoustic analysis generally indicated a lack of enhancement in IDS. Improved performance could be interpreted as evidence of contrast enhancement in both the clustering models and logistic regression. Regardless of the direction of these effects, the use of computational models provided an objective assessment of the discriminability of vowel categories in each register. This analysis also made explicit predictions about the generalisations that may be drawn from the statistical regularities in samples of IDS and ADS data. Comparisons of model performance therefore indicated the magnitude of the register-specific effects that were observed in the previous chapter. The acoustic analysis of IDS and ADS indicated the presence and direction of these effects but did not indicate whether cases of deterioration had a marginal effect or whether they severely hindered the use of statistical mechanisms in infancy.

These objective methods can also be viewed as a comment on the measures of discriminability that have been adopted in this domain. Chapters 3 and 4 presented measures of peripherality, dispersion and overlap in order to describe differences in the central tendency and variance of vowel categories across registers. The previous chapter favoured the degree of overlap between categories (D(a), Newman, Clouse, and Burnham, 2001) as a measure of discriminability since it could be related to the distributional properties of the input. The selection of a measure of discriminability has important implications for the interpretation of speakers' intentions in IDS. These measures define what constitutes evidence of a facilitative effect in IDS. If vowel space expansion does not adequately indicate greater discriminability, then this may invalidate studies that have adopted this measures and presented evidence of hyperarticulation in IDS. Methodological discussions of this type have been criticised for being dependent on researchers' intuitions about how the properties of the input benefit the infant learner (Eaves Jr. et al., 2016). Since these intuitions may be fallible, this critique proposes that computational models must be used in order to provide objective support for the use of certain measures of discriminability. The current chapter therefore considered the extent to which differences in model performance aligned with the results of the acoustic analysis from the previous chapter.

5.2 Methodology

This section describes the two modelling techniques that were applied to the acoustic data that was sampled from IDS and ADS. A clustering technique that used expectation maximisation replicated the inferential task that infants face in perceptual learning while a multinomial logistic regression assessed whether vowel quality distinctions were more easily discriminated in multidimensional acoustic space.

5.2.1 Materials

The speech corpus that was considered in the current analysis has been described in 3.1.1. The data consists of a series of unstructured interactions between four female speakers of American English and their infants as well as a series of directed interviews with an adult researcher. The first three formants, log vowel duration and patterns of spectral change were measured for each vowel to operationalise differences in quality across a set of fifteen vowel categories.

5.2.2 Clustering through Expectation Maximisation

Justification Though the acoustic analysis closely considered the distributional properties of the input that infants receive, this analysis did not indicate whether a statistical learner could successfully identify native language categories on the basis of these statistical regularities. The identification of an unknown set of categories within a given data set can be implemented as an unsupervised clustering task. Under the assumption that each vowel is a set of normal distributed tokens in acoustic space, a Gaussian mixture model may be used to describe the properties of the vowel inventory. Specifically, these models consist of a set of K probability distributions each defined by a component weight π_K, a mean μ_K, and a covariance matrix Σ_K.

Model specification For any given value of K, the parameters of each Gaussian that best fit the data can be located through the use of maximum likelihood estimation. Iterative techniques such as expectation maximisation (EM) are one method that has been used to identify an optimal set of parameters. Likelihood, $\hat{L}$, is maximised when each data point is minimally distant from the category that it is assigned to. Because of this, these models assigned tokens to categories on the basis of similarity. Though likelihood plays a major role in category assignment and parameter estimation, this statistic must be considered alongside a measure of the model's complexity in order to select an appropriate value for K. Clustering techniques, such as that implemented in the `R` package `mclust` (Fraley and Raftery, 2006), generate a series of models for a range of values of K and select the model with the most appropriate number of clusters using the Bayesian information criterion (BIC)[1].

$$\mathrm{BIC} = \ln(n)k - 2\ln(\hat{L})$$

The Bayesian information criterion is defined by an indicator of number of parameters that the model estimates, n, (i.e. values for π, μ and Σ for each category) minus the model's likelihood, $\hat{L}$. An appropriate value for K cannot be defined through complexity alone since this approach would favour models that consist of a single category, resulting in model underfit. Measures of likelihood penalise models that underfit the data: with a smaller number of categories, tokens would necessarily be assigned to high variance categories and reduce the model's likelihood. Likelihood cannot be used alone to selected values for K since this statistic is maximised when each data point is assigned to its own category, resulting in model overfit. The model with the lowest value for BIC is therefore selected as having the most appropriate number of clusters. Low values of BIC indicate low complexity, good fitness or an ideal trade-off between the two factors.

Data preparation In order to discover the types of inferences about category identity that could be drawn from the acoustic data, a separate clustering model was applied to speech data from each speaker and register. In addition to this, this analysis also considered how different sets of acoustic dimensions affected the number of clusters that were recovered for a given speaker and register. A minimal condition, where vowel quality was operationalised through measures of the first two formants, contrasted with a maximal condition that additionally considered F_3, vowel duration and the dynamics of F_1 and F_2 . Thus, sixteen sets of acoustic data were generated by resampling the original corpus data. This process of resampling allowed for the number of tokens and the relative frequency of each category to be controlled across the sixteen sets. For each speakers and register, I estimated the mean, covariance and frequency of each of the fifteen vowels. These estimates were used to generate a series of normal probability distributions that could be sampled in order to generate a set of 3000 vowel tokens that were representative of the original corpus data.

Separate clustering models were applied to each of these sets of data. The clustering models were implemented using the `Mclust` package in `R` and considered between 1 and 20 clusters (Fraley and Raftery, 2006). These models did not have any prior assumptions about the volume, shape or orientation of the clusters in the data. Since the estimation of the parameters of these models was probabilistic, a hundred runs of the model were carried out for each data set.

[1]BIC-based model selection stands in contrast to infinite Gaussian mixture models which infer K from the data (Feldman et al., 2013, , inter alia). These models are initialised with a potentially infinite number of categories and a bias towards smaller values for K. These models locate an appropriate number of categories through Markov chain Monte Carlo methods which eliminate clusters that do not further explain the data.

vowel	%	vowel	%	vowel	%
i	12.4	ɝ	5.9	ʊ	2.0
ɪ	11.1	ɑ	5.4	u	6.0
eɪ	5.8	ʌ	8.6	ɑɪ	8.2
ɛ	7.6	ɔ	4.4	aʊ	3.3
æ	11.1	oʊ	7.5	ɔɪ	0.7

Table 24: The frequency of each vowel category in the speech corpus, aggregated across speakers and registers.

Measures of model performance The outputs of this analysis were first analysed descriptively. The relative success of clustering techniques was judged by considering the number of categories the model selected. Infinite Gaussian mixture models of vowel systems that underfit the data have shown poor performance (Antetomaso et al., 2016; Feldman et al., 2013; Mooney, 2015) while the identification of a correct number of clusters has been used as a measure of success for other EM models (McMurray, Aslin, and Toscano, 2009; Vallabha et al., 2007). Further to this, the parameters of each cluster have also been examined as an indicator of success. Previous considerations of the learnability of IDS vowel categories (de Boer and Kuhl, 2003) have compared the outputs of an EM model with a fixed number of categories to the training data. In this case, models with parameters that closely resembled the input were judged to be successful.

In addition to this, model performance was also assessed through the use of pairwise F-scores. Measures of classification accuracy that depend on confusion matrices are of limited use when applied to models that must learn an unknown number of categories. In cases of overfit and underfit, it may be difficult to establish a set of correspondences between the categories in the input and those that are predicted by the model. Pairwise measures of accuracy were favoured in this instance since they are defined by considering of whether pairs of vowel tokens were judged to be members of the same category or not. True positives (TP) are cases where the model allocated a pair of tokens to the same class that actually were instances of the same vowel category. False positives (FP) occurred when the model considered pairs to be the same that should have been distinct while false negatives (FN) are pairs in the input that the model considered to be distinct. F-scores are defined as the harmonic mean of two accuracy measures, precision and recall.

$$\text{Precision} = \frac{N_{TP}}{N_{TP} + N_{FP}} \qquad \text{Recall} = \frac{N_{TP}}{N_{TP} + N_{FN}}$$

$$\text{F-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Reporting BIC probably isn't going to be a reliable way to determine model performance, except when considering different values for K. I'm also happy to expand this analysis by considering the Dunn index and such, but I think this is something we'd want to talk through in detail first.

5.2.3 Multinomial logistic regression

Justification The acoustic data did not conclusively address the claim that additional acoustic dimensions help to disambiguate the cases of overlap that have been reported for the distribution of the first two formants (Swingley, 2009). The previous analysis was only suggestive of this effect: a descriptive analysis of the third formant, patterns of spectral change and vowel duration associated these dimensions with category identity and further showed a numerical increase in dispersion for higher-dimensional space. A multinomial logistic regression was required in order to determine whether these additional dimensions provided reliable information about category identity in the current acoustic data sampled from IDS and ADS.

Model specification Logistic regressions indicate the extent to which a series of continuous variables affect the probability of responses to categorical variable. In general, a logistic regression considers a binary categorical response and outputs a series of coefficients that correspond to each independent variable. The estimate and standard error of each coefficient indicate whether the independent variables affect the categorical response: a Wald test can be used to determine whether each dimension significantly affected the probability of the categorical response. Rather than considering a single binary categorical response, multinomial logistic regressions select one category as a reference level and then implement a series of dummy variables for each other category in the system. Multinomial logistic regressions can therefore be viewed as a set of separate binary regressions for each dummy variable and thus estimate coefficients that determine how each acoustic dimension affects the selection of the reference category against all others. The identity of this reference category should therefore be considered when interpreting the results of these models.

Data preparation In order to explore how each acoustic dimension predicted category identity, multinomial logistic regressions were fitted to speech data from each speaker and register. These models predicted the identity of the fifteen vowels of American English on the basis of static measures of the first three formants, dynamic measures of F_1 and F_2, and measures of vowel duration. In order to facilitate this analysis, eight sets of acoustic data were generated by resampling the original corpus data. This process of resampling controlled for the differences in the total number of tokens and their relative frequencies. As above, I estimated the mean, covariance and frequency of each of the fifteen vowels for each speaker's IDS and ADS productions. These estimates were used to generate a series of normal probability distributions that could be sampled in order to generate a set of acoustic data that was representative of the original corpus data. Unlike the clustering approach above, the frequency of each phonemes was equated in this resampling process. Each set consisted 3000 acoustic tokens with 200 tokens of each of the fifteen categories of American English. The front open-mid unrounded vowel, /ɛ/, was selected as the reference category since this vowel had acoustic values that were close to the mean value for the majority of the acoustic dimensions that were considered in the model.

Measures of model performance The primary concern of this analysis was the set of coefficients that defined the extent to which each of the acoustic dimensions predicted differences in vowel quality. As stated above, the significance of the model coefficients were tested with a series of Wald tests. If acoustic dimensions beyond the first two formants contributed to vowel identity in the samples of IDS and ADS data, significant coefficients should be observed for F_3, vowel duration and the two measures of spectral change.

In addition to this, multinomial logistic regressions served as an alternative method of classifying the acoustic data. This supervised approach, where the model was trained on labeled data from the corpus, was contrasted with the clustering models. These regressions should be viewed as the performance of optimal observer, rather than as an approximation of the abilities of an infant learner. This supervised approach allowed for correspondences to be established between the actual categories in the corpus and those predicted by the model. Classification accuracy was therefore calculated by constructing a confusion matrix. These matrices were used to diagnose cases of overlap and to calculate measures of precision, recall and the F-score for each individual category. These cases of overlap were further considered with regard to the outputs of the clustering model, highlighting vowel distinctions that infants may fail to detect through the use of distributional learning. Unlike the pairwise measures of precision and recall, these measures pertained to individual categories rather than the system as a whole. True positives thus indicated cases where the model correctly classified tokens of a specific vowel. Precision indicated the percentage of selected tokens that were genuine members of the relevant category while recall indicated the percentage of the actual tokens that were predicted as such by the regression model.

5.3 EM clustering

5.3.1 Results

Number of clusters The viability of distributional learning was assessed by considering the number of clusters, K, that were selected as BIC-optimal across speakers and registers. Table 25 indicates the mean and standard deviation of the values of K that were selected across the 100 runs for each data set. Figures 40 and 41 indicate the distribution of these values of K for models applied to multidimensional data and formant data respectively. This analysis will first address models that received all of the available acoustic dimensions. In general, these models indicated that the distributional information in each register was not sufficient to recover the fifteen vowel

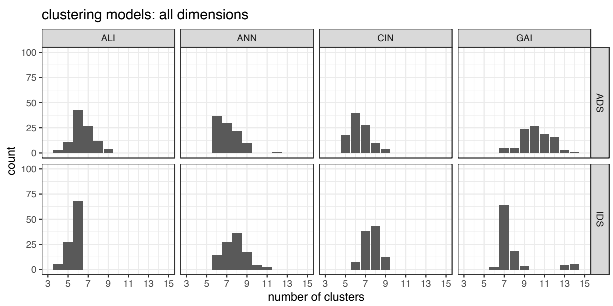

Figure 40: The distribution of K across the 100 models which were applied to multidimensional acoustic data across speakers and registers. Though K was less than fifteen in the majority of cases, the value of K was greater in ADS for ALI, CIN, and GAI while ANN had greater values for K in IDS.

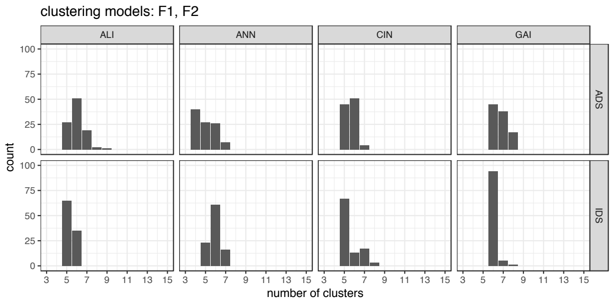

Figure 41: The distribution of K across the 100 models which were applied to formant distributions across speakers and registers. These models had lower values for K than those that were applied to multidimensional data, indicating that additional acoustic dimensions improved model performance.

	IDS		ADS	
all dims	mean	sd	mean	sd
ALI	5.63	0.58	6.46	1.07
ANN	7.76	1.14	7.10	1.11
CIN	7.60	0.79	6.42	1.02
GAI	8.13	2.36	10.32	1.36
F_1, F_2	mean	sd	mean	sd
ALI	5.35	0.48	5.99	0.80
ANN	5.93	0.62	5.00	0.97
CIN	5.56	0.88	5.59	0.57
GAI	6.07	0.29	6.72	0.74

Table 25: The number of clusters, K, selected on the basis of BIC for the vowel data across registers and speakers. The means and standard deviations reported here indicate model performance across 100 runs for each data set.

categories of American English. For the IDS vowel production data, the majority of the BIC-optimal models had between six and eight clusters. Models with values for K that approximated the actual number of categories in the system were rarely selected as optimal, indicating that greater complexity was not merited by an increase in model likelihood. In the minority of cases, models with thirteen or more clusters were optimal for speaker GAI's IDS production. For the ADS data, similar results obtained and the BIC-optimal models typically had between six and eight clusters for three of the four speakers. Models applied to GAI's ADS production data had highest mean value of K, 10.32. Underfit was still observed, however, since none of these models had fifteen clusters.

When models were fit to two-dimensional formant distributions, the BIC-optimal models had fewer categories than the high dimensional models. For the low-dimensional IDS data, the selected models typically had five or six clusters. Models with greater values for K were rarely selected and none of the optimal models had nine or more clusters. Similar results were seen for the models that were applied to the ADS data with typical solutions having between five and seven clusters. Though models with seven or eight clusters were selected for speaker GAI's data, these were infrequent and the value of K was consistently lower than when all acoustic dimensions were considered.

Model classifications Samples of the clusters that were identified through this method are presented in figures 42 and 43. The sample IDS models featured in these figures located BIC-optimal solutions that had either six or seven clusters. Each of these sample models identified two clusters with low values for F_1 which resembled /i/ and /u/. In addition to /i/, these models typically identified two clusters for the other front vowels in the data. One of these had a more close and central quality and consisted of tokens /ɪ/ and /ɝ/. The other had a more open quality and consisted tokens of /eɪ/, /ɛ/ and /æ/. These specific models also identified two clusters for back

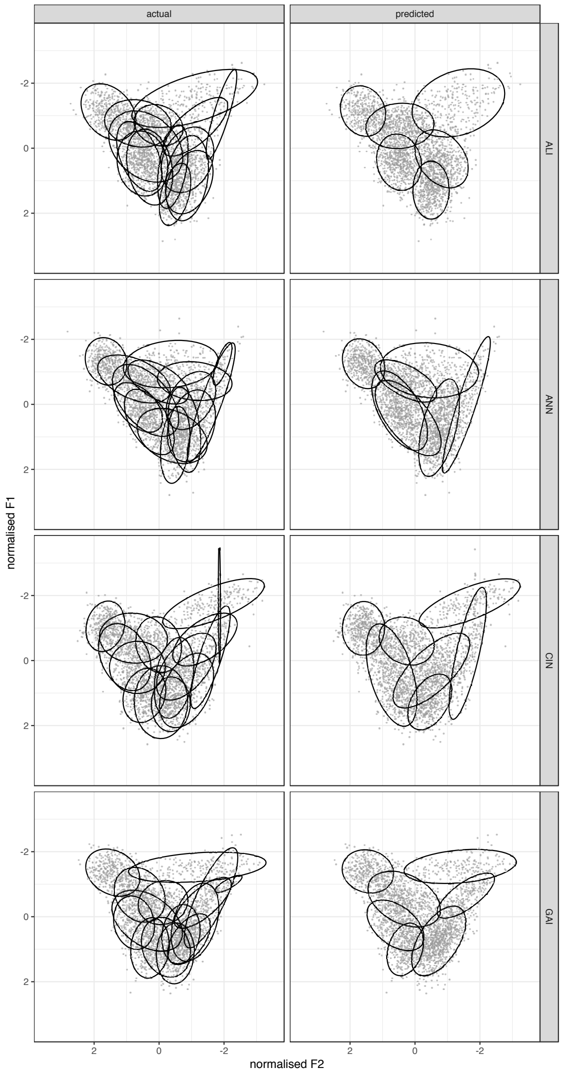

Figure 42: A comparison of actual IDS vowel categories (left) and those identified by BIC-based clustering models applied to multidimensional data (right). Each ellipse is a 95% confidence interval that corresponds to a category or cluster. This figure illustrates how the clustering models failed to detect relevant categories on the basis of distributional information

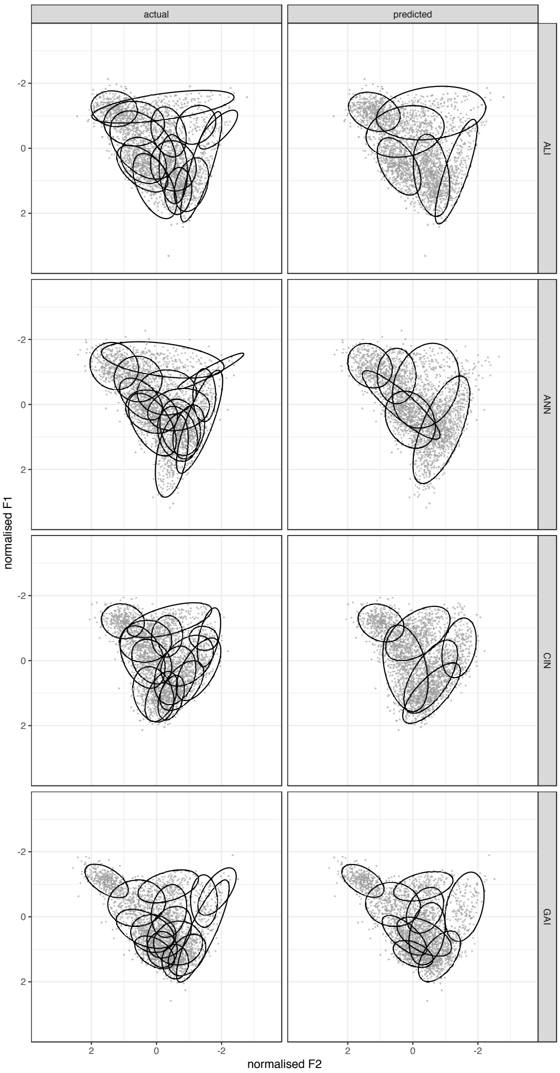

Figure 43: A comparison of actual ADS vowel categories (left) and those identified by BIC-based clustering models applied to multidimensional acoustic data (right).

vowels other than /u/. Models applied to ALI and ANN's productions had one mid cluster that consisted of tokens of /ɔ/ and /oʊ/ and a more open cluster which consisted of tokens of /ɑ/, /ʌ/ and /ɑɪ/. By contrast, models applied to CIN and GAI's data identified a cluster which spanned the entire range of F_1 and consisted of tokens with a minimal value for F_2 and a further cluster (or pair of clusters) which consisted of tokens of /ɑ/, /ʌ/, /oʊ/, /ɔ/, and /ɑɪ/.

Similar results were observed for the models applied to speakers' ADS production data. The sample models applied to vowels produced by ALI, ANN and CIN in this register located six categories while the sample model for GAI located ten clusters. Each of these sample models again identified a close, front cluster that corresponded to the vowel /i/. A corresponding category for /u/ was only observed in the models applied to the productions of ALI and GAI. The sample model for ALI had two front vowel clusters front vowels which separated /ɪ/ and /ɝ/ from /eɪ/, /ɛ/ and /æ/. Two back clusters were also identified, separating /aɪ/, /aʊ/, and /ʌ/ from /ɔ/ and /ɑ/. The model applied to ANN's data had three front clusters: one of these aligned well with /I/ and the other two overlapped considerably and divided tokens of /eɪ/, /ɛ/ and /æ/ between them. A central category consisted of tokens of /ɝ/, /ʊ/ and /u/ and a final cluster consisted of all of the low back vowels. The model applied to CIN's data had a central cluster which consisted of tokens /ɪ/, /ɝ/, and /u/ and a front cluster that consisted of every front vowel other than /i/. One back cluster consisted of tokens of /oʊ/, /ʊ/, and /u/ while the low back vowels were assigned to one of two overlapping clusters. The model applied to GAI's data showed the strongest correspondence to the category structure of the input. Five of the clusters corresponded to /i/, /u/, /ɪ/, /ɝ/, and /oʊ/. Three overlapping categories showed poorer correspondence with the remaining front vowels while two clusters consisted of back vowels.

Pairwise accuracy scores The accuracy of the clustering techniques was assessed with pairwise measures of precision and recall and their harmonic mean, the F-score. The accuracy measures reported in tables 26 and 27 indicated that model performance was poor and were consistent with the observation of patterns of underfit. Low values for pairwise precision indicated that the clustering models had a high number of false positives. Thus, these models assigned many tokens that were actually distinct to a single cluster. The comparatively high values for pairwise precision were also consistent with underfit. Since the models posited a small number of clusters and thus did not consider many pairs to be different, there were comparatively few false negatives. The highest values for the F-score were observed in models that were fit to speaker GAI's ADS data: the BIC-optimal models for this data had larger values for K than any of the others. Though there was a relative improvement of accuracy, measures of pairwise precision indicated that many relevant quality distinctions were not detected. The values for pairwise recall also should be interpreted cautiously: though these models had the greatest number of clusters, they still underfit the data and thus high values for recall are expected.

IDS, all dimensions				
	F-score	Precision	Recall	*K*
ALI	.396 [.391, .401]	.276 [.271, .282]	.708 [.700, .716]	5.63
ANN	.485 [.479, .491]	.374 [.367, .382]	.694 [.689, .698]	7.76
CIN	.441 [.436, .447]	.323 [.316, .330]	.708 [.699, .716]	7.60
GAI	.476 [.473, .479]	.371 [.363, .378]	.682 [.669, .694]	8.13
ADS, all dimensions				
	F-score	Precision	Recall	*K*
ALI	.464 [.460, .469]	.353 [.347, .359]	.683 [.679, .687]	6.46
ANN	.436 [.430, .442]	.327 [.321, .334]	.658 [.655, .662]	7.10
CIN	.459 [.452, .465]	.340 [.332, .349]	.714 [.709, .719]	6.42
GAI	.608 [.602, .615]	.532 [.521, .543]	.718 [.713, .723]	10.32

Table 26: Pairwise F-scores, precision and recall with 95% confidence intervals for the clustering models applied to multidimensional data across speakers and registers. Values for the pairwise F-score and precision indicate that models fit too few clusters to the data. Values for pairwise recall indicated that these models generally assigned tokens of the same vowel category to the same cluster.

IDS, F_1, F_2				
	F-score	Precision	Recall	*K*
ALI	.388 [.384, .391]	.266 [.262, .271]	.722 [.713, .731]	5.63
ANN	.399 [.396, .401]	.284 [.280, .288]	.675 [.668, .681]	5.93
CIN	.401 [.396, .407]	.280 [.273, .286]	.724 [.717, .731]	5.56
GAI	.407 [.403, .411]	.289 [.286, .293]	.685 [.677, .692]	6.07
ADS, F_1, F_2				
	F-score	Precision	Recall	*K*
ALI	.409 [.406, .413]	.292 [.287, .296]	.695 [.688, .701]	5.99
ANN	.407 [.403, .412]	.288 [.283, .294]	.700 [.691, .709]	5.00
CIN	.402 [.399, .405]	.284 [.280, .287]	.692 [.686, .698]	5.59
GAI	.426 [.420, .431]	.312 [.303, .321]	.682 [.672, .692]	6.72

Table 27: Pairwise F-scores, precision and recall with 95% confidence intervals for the clustering models applied to formant distributions across speakers and registers.

Differences in the pairwise F-scores of these clustering models were considered in two follow-up analyses. The first considered how register-specific differences in vowel production affected model performance while the second considered the inclusion of acoustic cues beyond the first two formants. F-scores from across the 100 model runs were compared across registers using a series of T-tests. When all possible acoustic cues were considered, higher F-scores were observed for the ADS models compared to the IDS models for production data from speakers ALI ($t(197) = 10.98$, $p < .001$), CIN ($t(184) = 4.00$, $p < .001$) and GAI ($t(138) = 35.07$, $p < .001$). Conversely, clustering models showed improved performance when applied to speaker ANN's ADS productions in comparison to her IDS data ($t(197) = -11.8$, $p < .001$).

For models applied to the distribution of the first two formants, improved performance was seen for ADS data in comparison to IDS data for three of the four speakers (ALI, $t(197) = 19.41$, $p < .001$; ANN $t(114) = 7.47$, $p < .001$; GAI, $t(195) = 7.18$, $p < .001$). The performance of these models did not differ across registers when they were applied to speaker CIN's data ($t(152) = 0.06$, $p = .949$).

The second follow-up analysis demonstrated that model performance improved when this clustering technique was applied to high dimensional acoustic data in comparison to when it was applied to to formant distributions. A series of T-tests indicated that pairwise F-scores were greater in the high dimensional data extracted from the IDS productions of each of the four speakers (ALI, $t(175) = 2.76$, $p < .001$; ANN, $t(136) = 26.17$, $p < .001$; CIN $t(197) = 10.67$, $p < .001$; GAI, $t(185) = 27.02$, $p < .001$). A similar set of analyses also showed that models applied to high dimensional data sampled from ADS outperformed those that were applied to IDS. This positive effect of high dimensional data was observed for all four speakers (ALI, $t(170) = 19.24$, $p < .001$; ANN, $t(160) = 13.38$, $p < .001$; CIN $t(134) = 15.25$, $p < .001$; GAI $t(172) = 43.48$, $p < .001$).

5.3.2 Interim discussion

The current set of clustering models aimed explore whether the vowel quality distinctions of American English could be recovered on the basis of the statistical regularities that occurred the acoustic data that was sampled from IDS and ADS. Model underfit was consistently observed for both registers as these models generally identified between six to eight clusters rather than the fifteen vowels of American English. This analysis indicated that individual modes for each phonemic category could not be located in the frequency distributions of the multidimensional acoustic data. As these models closely replicated distributional learning in infancy, these results indicated that a statistical learner exposed to input from either register would collapse native language distinctions rather than preserving them. Though some clusters did correspond to the peripheral vowels /i/ and /u/, other clusters consisted of tokens of two or more native language categories. Models with fifteen clusters were not selected as optimal as they had low values for the BIC, indicating that they did not show the increase in likelihood that was required to justify their additional complexity.

The poor model performance that was observed across speakers and registers was consistent with the acoustic analysis of IDS and ADS. Visual inspections of the formant distributions in the data indicated a considerable degree of category overlap in both registers. The current set of results were comparable to previous studies which have applied to clustering models to vowel production data and thus suggested that distributional learning cannot be successfully applied to entire inventories. Similar cases of model underfit have been observed for American English (Feldman et al., 2013) and Scottish English ADS data (Mooney, 2015). Models of American English and Japanese IDS data have conversely demonstrated cases of model overfit (Antetomaso et al., 2016). Unsupervised clustering models have recovered an appropriate number of categories when only a subset of the vowel distinctions in a language were considered (Dutch (/ɑ/, /aː/: Benders, 2013; French /i/, /y/, /u/: Moeng, 2016; Japanese /i/, /iː/; /ɛ/, /ɛː/, American English /i/, /ɪ/; /eɪ/, /ɛ/: Vallabha et al., 2007). The improved model performance in these subsets with respect to entire inventories can be interpreted in two ways. One possibility is that models that consider a subset of distinctions have presented unrealistic simplifications of the infant learning task. By excluding confusable alternatives, these models have overstated the viability of distributional learning in infancy. Another alternative is that distributional learning only enables infants to identify relevant categories within small samples of acoustic data and that this mechanism is inappropriate for for determining the number of categories within a language's inventory as a whole.

These clustering models extend previous assessments of the viability of distributional learning by considering high dimensional data from both IDS and ADS. Previous models may have shown poor performance by failing to account for register-specific differences in vowel production and/or the influence of additional acoustic cues. Comparisons of model performance across registers present a novel contribution to the field. To my knowledge, unsupervised clustering models with an unknown number of categories have only ever applied to acoustic data from a single register. The results of these clustering techniques indicated the discriminability of contrasts in each register and thus provided further insights into the claim that IDS vowels are more discriminable than their ADS counterparts. This analysis did not provide evidence of enhancement in IDS as differences in model performance aligned with the the differences in overlap that were observed across registers. This analysis therefore suggested that measures of overlap are a relevant measure of discriminability since the affected the performance of with models that directly replicate distributional learning in infancy.

The current set of models showed an improvement in performance where they were applied to high dimensional data in comparison to formant data sampled from each speaker and register. Because of this, these clustering models supported the claim that acoustic dimensions beyond the first two formants provide learner with relevant information about native language distinctions and thus mitigate the ambiguity of the input (Swingley, 2009). These results aligned with previous studies that have applied clustering models to multidimensional data. The use of multidimensional acoustic data enables models to identify a distinction between Dutch /ɑ/ and /aː/ (Benders, 2013) as well as a covert voicing contrast in Dutch final stops (Kirby, 2014). Unlike these models, however, the consideration of multidimensional data did not allow for an appropriate number of categories to be learnt from the current samples of American English IDS and ADS. Cases of overlap were mitigated in high dimensional space but not obviated in their entirety.

In summary, a series of clustering models were unable to recover the vowel inventory of American English by exploiting the statistical regularities that were apparent in samples of multidimensional acoustic data from IDS and ADS. This analysis suggested that distributional learning cannot be the sole explanatory mechanism behind perceptual learning in infancy. The distributional information in naturalistic samples of linguistic input did not allow learners to identify the number of categories that exist in American English or the parameters that define them. The clustering models that were presented in this chapter extended previous work in two ways. Firstly, comparisons of the performance of IDS and ADS models aligned with the register-specific differences in overlap that were observed in the acoustic analysis of these registers. This meant that model performance did not improve as a result of hyperarticulation in IDS. Models that were applied to multidimensional acoustic data also showed improved performance over models applied to formant distributions. This indicated that the poor performance of previous models could partially be attributed to a failure to consider a broad enough set of dimensions. That said, the fact that performance was poor across all of the clustering models presented here brings the relevance of these factors into question. Even in optimal conditions, these results predict that exposure to naturalistic input would lead statistical learners to collapse native language distinctions rather than maintain them.

5.4 Multinomial logistic regressions

5.4.1 Results

	IDS	ADS
ALI	.639	.732
ANN	.694	.664
CIN	.708	.714
GAI	.707	.773

Table 28: Measures of classification accuracy for the multinomial regressions applied to each data set

Classification accuracy Multinomial logistic regressions were applied to IDS and ADS speech corpora in order to assess the relative contribution of each acoustic dimension. In addition to this, the outputs of these models gave further insights into the status of individual distinctions in the input through category-specific measures of precision and recall. Unlike the EM clustering methods, this method should not be viewed as an attempt to replicate the infant learning process. Instead, this supervised approach indicated the inferences that an ideal observer may make about vowel quality distinctions in each of the samples of acoustic data.

Table 28 indicates how measures of classification accuracy differed across each speaker and register. Since this model was trained on a specific number of categories, correspondences could be established between the actual identity of vowels in the data and the model's predictions unlike in the EM cluster models. Measures of classification accuracy were calculated using the set of confusion matrices in figure 44 for IDS and

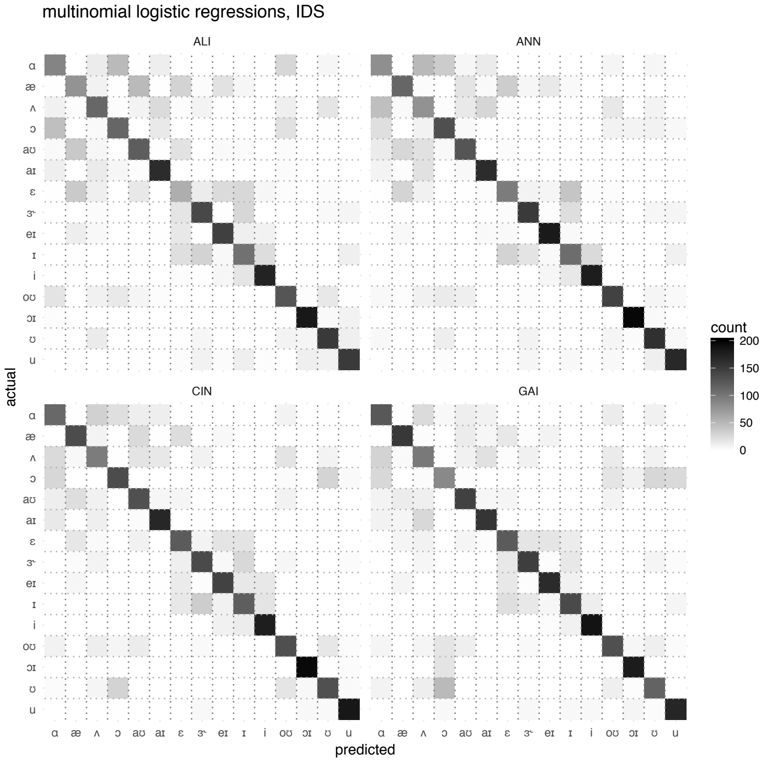

Figure 44: Confusion matrixes which indicate the alignment between the predictions of multinomial logistic regressions for each speaker's IDS data and the actual identity of vowels. Darker shading indicates that the model made the relevant prediction a greater number of times.

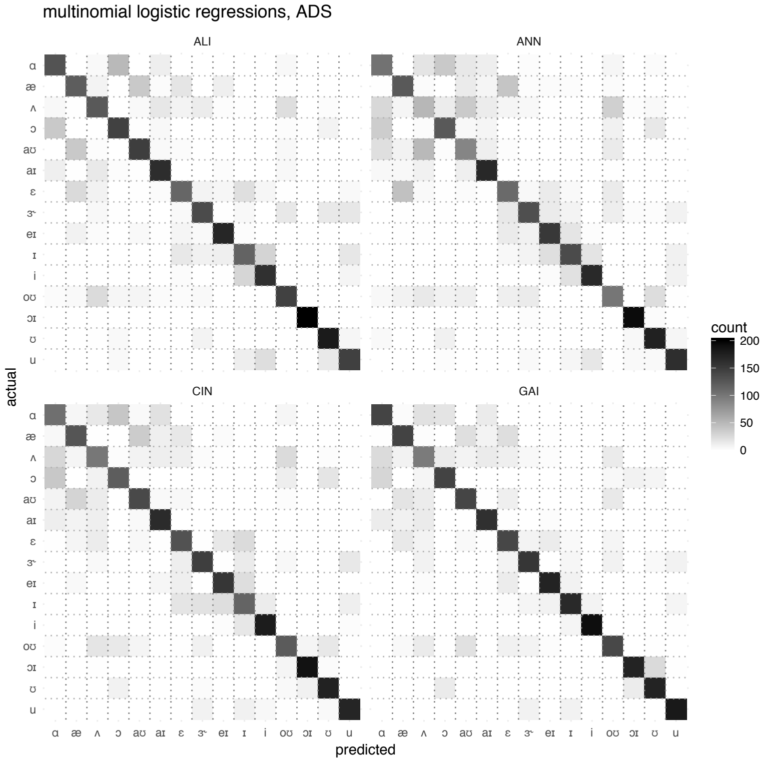

Figure 45: Confusion matrixes which indicate the alignment between the predictions of multinomial logistic regressions for each speaker's ADS data and the actual identity of vowels. Darker shading indicates that the model made the relevant prediction a greater number of times.

figure 45 for ADS. Versions of these confusion matrices which have actual numerical values that provide further detail concerning how the predicted and actual categories aligned are available in appendix A. In addition to measures of classification accuracy, model performance was assessed for individual categories through measures of precision, recall and F-scores. The confu- sion matrices were also inspected visually to identify cases where misclassification was common, indicating that the data showed a considerable degree of category overlap.

The scores reported for classification accuracy indicated that vowels were easier to discriminate in ADS for speakers ALI, CIN and GAI. Conversely, models showed improved performance in IDS for speaker ANN. These results were parallel to the differences in the pairwise F-score of the EM clustering models: in both of these cases, register-specific differences in model performance aligned with the degree of contrast overlap of IDS and ADS, as defined by measures of D(a).

Confusion matrices and category F-scores The confusion matrices in figures 44 and 45 were further used to calculate the F-scores for each category. Category-specific measures of the F-score are displayed in figure 46. The highest F-scores, indicating good model performance, were seen for the diphthong /ɔɪ/ and two peripheral vowels, /i/ and /u/. The diphthong had an F-score greater than .8 in each of the eight logistic regressions while /i/ and /u/ reached this criterion in seven and six models respectively. Models were expected to reliably identify /i/ and /u/ given that these are point vowels with extreme values for F_1, F_2 and F_3. The diphthong /ɔɪ/ also had low values for the first two formants. In addition to this, this vowel had a larger change in F_2 than any other vowel in the system.

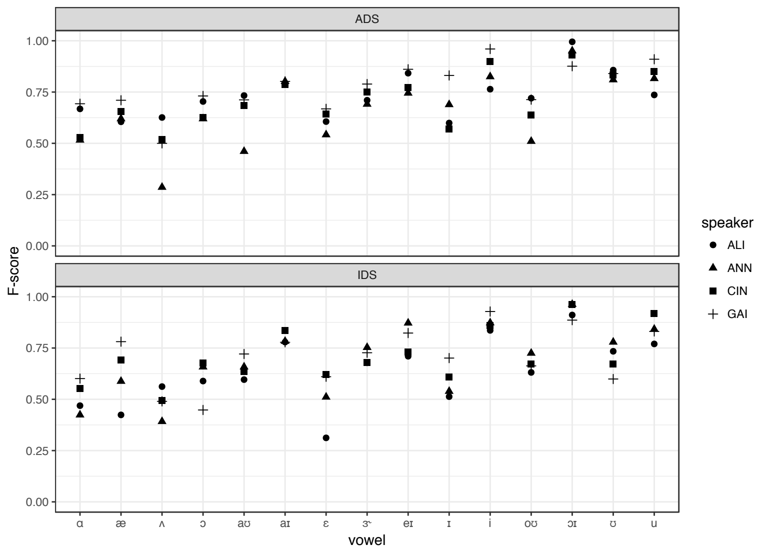

Figure 46: F-scores for each vowel category in the multinomial logistic regressions that were run for each register and speaker. This statistic is defined as the harmonic mean of precision and recall.

Cases of poor model performance, indicated by lower values for the F-score, were examined further in order to provide a qualitative assessment of the errors that the model made. Specifically, the current descriptive analysis considered confusable alternatives for each vowel that had a recall of .6 or less. The vowels /ɛ/ and /ɑ/ had a low value for recall in five of the eight models. In the majority of cases, instances of the open-mid front unrounded vowel /ɛ/ were mislabeled as /æ/, /ɪ/ or /eɪ/. The low back unrounded vowel /ɑ/ was frequently mislabeled as /ɔ/, /ʌ/ or /ɑɪ/. Pairs of vowels were frequently mislabeled as each other: the open-mid back unrounded vowel /ʌ/ had a recall of .6 or less in four models and was commonly identified as /ɑ/. Both /ɪ/ and /æ/ had a recall of .6 or less in two of the eight models. The vowel /ɪ/ was commonly labeled as /i/ or /ɛ/ while /æ/ was labeled as either /ɛ/ or /aʊ/.

F_1 The regressions indicated that the first formant which is associated with vowel height was a strong predictor for category identity in both IDS and ADS. For IDS vowels, the regression coefficients indicated that F_1 was a strong predictor of vowel identity for all categories in the models for the data produced by ALI and ANN. For CIN, the model coefficients were significant for all vowels other than /oʊ/ while both /oʊ/ and /eɪ/ failed to reach significance in the model of GAI's distinctions. Comparable results were seen in the analysis of the ADS production data. For speaker ALI, F_1 was a significant predictor of identity for all vowels other than /ɔɪ/ while coefficients for /eɪ/ and /oʊ/ failed to reach significance in speaker CIN's model. F_1 was a strong predictor of identity for all categories in models applied to the production of ANN and GAI. The outcomes of these models were consistent with the reference category that was selected: coefficients only failed to reach significance when /ɛ/ was contrasted with other mid vowels.

F_2 The second formant, associated with vowel backness, was also a strong predictor for category identity. For IDS vowels, the regression coefficients indicated that F_2 was a strong predictor of vowel identity for all categories in the models for distinctions produced by ALI, CIN and GAI. For speaker ANN, only the coefficient for /ɪ/ failed to reach significance. Regressions applied to the ADS production data also indicate that the second formant was strong predictor of category identity. Coefficients were significant for every vowel category in the models applied to ALI and GAI's production data. Similar results were seen in the majority of cases in the models applied to ANN and CIN's productions. Regression coefficients only failed to reach significance for the vowel /æ/ in both of these models. Again, the choice of the reference category should be considered here: the vowels /ɛ/ and /æ/ are both low front vowels with a similar degree of advancement.

F_3 Multinomial logistic regressions indicated that F_3 was a less reliable predictor of category identity than the first two formants. The vowel categories that had significant coefficients for the third formant are indicated in table 29. Models applied to IDS vowel data produced by ALI, ANN and CIN only had significant coefficients for F_3 for a minority of vowels. For speaker ALI, F_3 was predictive of identity for six vowels. This dimension predicted identity for four vowels in ANN's model and five in CIN's. This dimension was a strong predictor in GAI's IDS data: coefficients for F_3 only failed to reach significance for /ʊ/ and /ɔɪ/. A summary of the vowels that had significant coefficients is provided in table 29 which also details the outputs of the ADS models.

The models applied to the ADS production data indicated that F_3 was predictive of category identity for a larger number of vowels than those applied to the IDS data. ALI and ANN respectively had significant coefficients for eleven and ten of the vowels in the system. This dimension was a poorer predictor for CIN and GAI's production data with only seven vowels having significant coefficients: four of these vowels were common across the two speakers, namely /ɑ/, /ɔ/, /ʊ/, and /ɔɪ/.

speaker	IDS	ADS
ALI	i, eɪ, ɔ, ʊ, u, aʊ	i, ɪ, eɪ, æ, ɑ, ʌ, ɔ, oʊ, u, aʊ, ɑɪ
ANN	ɔ, ɝ, ɑɪ, ɔɪ	ɪ, eɪ, æ, ɝ, ʌ, ɔ, ʊ, u, aʊ, ɔɪ
CIN	eɪ, ʌ, ɔ, ʊ, ɔɪ	eɪ, ɑ, ʌ, ɔ, ʊ, ɑɪ, ɔɪ
GAI	i, ɪ, eɪ, æ, ɝ, ɑ, ʌ, ɔ, oʊ, u, aʊ, ɑɪ	ɝ, ɑ, ɔ, oʊ, ʊ, u, ɔɪ

Table 29: Vowels which had a significant coefficient for F_3 in the multinomial logistic regressions for each register and speaker.

Log duration In the models applied to the IDS data, the predictiveness of vowel duration varied across speakers. The vowel categories that had significant coefficients for log duration are indicated in table 30. The models applied to ALI and ANN's data had significance coefficients for duration for eight and nine vowels respectively. Duration was a poorer predictor for CIN's distinctions as only the coefficients for four vowels – /eɪ/, /ʌ/, /oʊ/, and /aʊ/ – were significant. Conversely, vowel duration was a strong predictor for all but two vowels in GAI's data: coefficients for the lax vowels /ɪ/ and /ʌ/ did not reach significance.

The models applied to the ADS data indicated that vowel duration was a consistent predictor of differences in vowel quality in this register. For ALI, all vowels other than /i/, /ɪ/ and /ɑɪ/ had significant coefficients for duration. Duration was least reliable for speaker ANN's data with six vowels failing to reach significance. Coefficients were significant in the majority of cases for speakers CIN and GAI with only three and four vowels failing to reach significance respectively. Across the four speakers, the identity of eight vowels was consistently predicted by their duration – /eɪ/, /æ/, /ɝ/, /ɔ/, /oʊ/, /u/, /aʊ/ and /ɔɪ/ had significant coefficients in all of the ADS models. It should be noted that /ɛ/ had a short vowel duration relative to other vowels across all four speakers in ADS, potentially explaining the stronger predictive effect seen in this register.

speaker	IDS	ADS
ALI	ɪ, eɪ, ɝ, ʌ, oʊ, u, ɑɪ, ɔɪ	eɪ, æ, ɝ, ɑ, ʌ, ɔ, oʊ, ʊ, u, aʊ, ɔɪ
ANN	eɪ, æ, ɝ, ɔ, oʊ, ʊ, u, aʊ, ɑɪ,	eɪ, æ, ɝ, ɔ, oʊ, ʊ, u, aʊ
CIN	eɪ, ʌ, oʊ, aʊ	eɪ, æ, ɝ, ɑ, ɔ, oʊ, ʊ, u, aʊ, ɔɪ
GAI	i, eɪ, æ, ɝ, ɑ, ɔ, oʊ, ʊ, u, aʊ, ɑɪ, ɔɪ	ɪ, eɪ, æ, ɝ, ɑ, ɔ, oʊ, u, aʊ, ɑɪ, ɔɪ

Table 30: Vowels which had a significant coefficient for log duration in the multinomial logistic regressions for each register and speaker.

Change in F_1 Patterns of spectral change capture the dynamics of the first and second formants throughout a vowel's duration. The multinomial logistic regression indicated that changes in the first formant were a strong predictor of category identity. The vowel categories that had significant coefficients for the change in the first formant are indicated in table 31. The poorest predictive capabilities were seen in ALI and GAI who had significant coefficients for all but four of the vowels in the system. Change in F_1 was a predictor of identity for all vowels other than /ɔ/ while CIN had significant coefficients for all vowels other than /ɑ/ and /aʊ/.

Similar performance was seen when these models were applied to the ADS data. The model for speaker ALI's data had significant coefficients for all vowels other than /aɪ/. For ANN, coefficients were significant for all vowels other than /ʌ/ and /aʊ/. This dimension was also a strong predictor of identity for speakers CIN (all except /æ/, /ʌ/, and /u/) and GAI (all except /i/, /æ/, and /ʌ/).

speaker	IDS	ADS
ALI	i, eɪ, æ, ɝ, ʌ, oʊ, ʊ, u, aʊ, ɑɪ	i, ɪ, eɪ, æ, ɝ, ɑ, ʌ, oʊ, ɔ, ʊ, u, aʊ, ɔɪ
ANN	i, ɪ, eɪ, æ, ɝ, ɑ, ʌ, oʊ, ʊ, u, aʊ, ɑɪ, ɔɪ	i, ɪ, eɪ, æ, ɝ, ɑ, oʊ, ɔ, ʊ, u, ɑɪ, ɔɪ
CIN	i, ɪ, eɪ, æ, ɝ, ʌ, oʊ, ɔ, ʊ, u, ɑɪ, ɔɪ	i, ɪ, eɪ, ɝ, ɑ, oʊ, ɔ, ʊ, u, aʊ, ɑɪ, ɔɪ
GAI	i, ɪ, eɪ, ɝ, ɑ, ʌ, oʊ, u, aʊ, ɑɪ	ɪ, eɪ, ɝ, ʌ, oʊ, ɔ, ʊ, u, aʊ, ɑɪ, ɔɪ

Table 31: Vowels which had a significant coefficient for change in F_1 in the multinomial logistic regressions for each register and speaker.

Change in F_2 Changes in the second formant were also a reliable predictor of vowel quality, as indicated in table 32. In the models applied to the IDS data, this dimension showed the poorest predictive capabilities for speaker ANN relative to other speakers. All but four vowels had significant coefficients for this acoustic dimension. For ALI and CIN, coefficients only failed to reach significance for three vowels in the system while the change in F_2 in GAI's data was predictive of identity for all vowels other than /oʊ/ and /ɑ/.

For the ADS data, the change in F_2 showed comparable performance. The relative predictability of this dimension across speakers was poorest for ANN: the model coefficients for this speaker were significant for nine of the fourteen distinctions that were considered. ALI had significant coefficients for all vowels other than /ɔ/ while this dimension was a predictor of twelve of the fourteen vowels for speakers CIN (all except /ʌ/ and /ʊ/) and GAI (all except /æ/ and /ʌ/).

speaker	IDS	ADS
ALI	i, eɪ, æ, ɝ, ʌ, oʊ, ɔ, u, aʊ, ɑɪ, ɔɪ	i, ɪ, eɪ, æ, ɝ, ɑ, ʌ, oʊ, ʊ, u, aʊ, ɑɪ, ɔɪ
ANN	i, eɪ, ɝ, ɑ, oʊ, ʊ, u, aʊ, ɑɪ, ɔɪ	i, ɪ, eɪ, ɝ, oʊ, ʊ, u, aʊ, ɑɪ, ɔɪ
CIN	i, eɪ, ɝ, ɑ, ʌ, oʊ, ʊ, u, aʊ, ɑɪ, ɔɪ	i, ɪ, eɪ, æ, ɝ, ɑ, oʊ, ɔ, u, aʊ, ɑɪ, ɔɪ
GAI	i, ɪ, eɪ, æ, ɝ, ʌ, ɔ, ʊ, u, aʊ, ɑɪ, ɔɪ	i, ɪ, eɪ, ɝ, ɑ, oʊ, ɔ, ʊ, u, aʊ, ɑɪ, ɔɪ

Table 32: Vowels which had a significant coefficient for change in F_2 in the multinomial logistic regressions for each register and speaker.

5.4.2 Interim discussion

The multinomial logistic regressions that were presented in this chapter aimed to demonstrate how dimensions beyond F_1 and F_2 contributed to vowel quality in IDS and ADS. Further to this, they aimed to describe the performance of an optimal observer who is fully capable of exploiting the distributional properties of the input. As expected, these logistic regressions indicated that the first two formants were predictive of category identity for the majority of vowels across speakers and registers. Significant regression coefficients were also observed for duration and patterns of spectral change for the majority of vowel categories in each model. Coefficients for F_3, however, were significant for a smaller set of vowels and thus indicated that this dimension was a less reliable predictor of category identity. These results therefore supported the claim that the ambiguity in the acoustic input may be mitigated by considering these additional acoustic dimensions (Swingley, 2009). The acoustic analyses in chapter 4 only provided partial evidence for this claim in the form of the numerical increases in Euclidean distance and D(a) that were observed in multidimensional acoustic space. Though the improved performance of clustering analysis that were applied to high dimensional data was consistent with this effect, this analysis did not indicate how each acoustic dimension contributed to specific distinctions in the vowel system of American English. Measures of classification accuracy of the models that were applied to IDS and ADS indicated that both of these registers present the learner with ambiguous distributional information with respect to vowel identity in American English. These results demonstrated that even an optimal observer will assign vowel tokens to the incorrect category if they solely consider the acoustic properties that were described in this analysis.

The current use of logistic regressions extended previous studies in this domain that adopted this technique. This type of model has been used to identify whether specific dimensions provide the learner with reliable distributional information for a given distinction. Such analyses have demonstrated that vowel duration was a strong predictor of phonological length in Japanese IDS (Bion et al., 2013). Similar models have indicated that the acoustic properties of IDS provides learners with reliable predictor for a pair of comparable distinctions in Japanese (/i, iː; ɛ, ɛː/) and American English (/i, ɪ; eɪ, ɛ/; Werker et al., 2007). The current analysis extended this work by assessing how a broader set of acoustic dimensions predicted the identity of the full set of vowel categories in American English. The regressions presented in this chapter indicated that dimensions beyond the first two formants were predictive of identity in the vowel production data that was sampled from the IDS and ADS corpora. Additionally, these regressions did not suggest that the acoustic dimensions in IDS were reliable predictors of identity. The larger number of distinctions that was considered in the current analysis may explain this differences in the predictability of categories as each vowel was considered in the context of multiple confusable alternatives rather than as a member of a single distinction.

Further comparisons can be drawn between the current set of models and a regression analysis that was presented in support for a comparative acoustic analysis of American English vowels in IDS and ADS (McMurray et al., 2013). These regressions considered the extent to which F_1, F_2 and F_3 predicted the identity of eight vowels (/i/, /eɪ/, /æ/, /ɝ/, /ɑ/, /ʌ/, /oʊ/, and /aɪ/) in each register. Both this previous analysis and the one presented in this chapter had values for classification accuracy which indicated that the acoustic properties of both IDS and ADS presented ambiguous cues to category identity. The analysis of eight vowels also indicated that classification accuracy was greater in ADS than IDS and viewed this result as evidence that the acoustic properties of IDS did not support the identification of native language distinctions in infancy. Both of these analyses used measures of accuracy to report individual differences in the discriminability of specific vowel categories. (McMurray et al., 2013) observed that peripheral vowels such as /i/, /æ/, and /oʊ/ had higher accuracy scores than interior vowels such as /ɪ/, /ɝ/, and /ʌ/. The results of the current analysis aligned partially with these observations as vowels with extreme acoustic properties such as /i/, /u/, and /ɔɪ/ had greater accuracy than interior vowels such as /ɛ/, /ʌ/ and /ɪ/. Unlike the previous analysis of American English, the accuracy of /oʊ/ and /æ/ was not at ceiling in the current analysis. Again, the exclusion of confusable alternatives may have resulted in a overstatement of the discriminability of vowels: for example, /oʊ/ may have been less discriminable in the current study as it was considered alongside /ɔ/, /ʊ/ and /u/.

In summary, logistic regressions indicated that log duration and patterns of spectral change were strong predictors of vowel identity in the vowel production data that was reported here. The third formant was a less reliable predictor of category identity across the system as a whole. The use of regressions to assess the extent to which multiple acoustic dimensions predicted each of the vowel categories of American English was a novel contribution of the current study. These models supported the claim additional cues may facilitate the identification of vowel categories in infancy by reducing the ambiguity of the input that learners receive (Swingley, 2009). Despite this, the use of higher dimensional data did render the infant learning task trivial. The fact that measures of classification accuracy were not at ceiling for either IDS and ADS indicated that the acoustic input was ambiguous. This ambiguity suggested that distributional learning was not sufficient to explain perceptual attunement in infancy as the capabilities of this supervised learner with capabilities that cannot be attributed to infant learners still made errors in categorisation.

5.5 General discussion

In contrast to the acoustic analyses which primarily consider the extent to which the acoustic properties of IDS were consistent with contrast enhancement in IDS, the current chapter presented a series of computational models which assessed the viability of distributional learning in infancy. Experimental paradigms have demonstrated that infants track the statistical regularities of the acoustic input and that their perceptual behaviour becomes aligned with what they observe (Maye, Werker, and Gerken, 2002). Because of this, distributional learning has been viewed as a potential explanatory mechanism for perceptual attunement in infancy. The use of this mechanism outside of laboratory contexts depends on an assumption that the input that infants receive presents them with reliable information about category identity. The two sets of models presented in this chapter demonstrated that the statistical regularities of the input did not allow for the recovery of a set of categories which corresponded to the fifteen phonemes of American English. Clustering models consistently collapsed native language distinctions and the logistic regressions made classification errors despite the supervised nature of this approach.

The performance of these models is summarised in table 33. This table indicates that the number of categories, K, which the clustering models located was smaller than the number of phonemic vowels of American English. Further to this, this table indicates how the pairwise F-score of these models differed across registers and as a function of the number of acoustic dimensions that these models were included in the input. This table further reports the classification accuracy of the logistic regressions, indicating that model performance was not at ceiling for either register.

	ALI	ANN	CIN	GAI
mean K, IDS, all dimensions	5.63	7.76	7.60	8.13
ADS, all dimensions	6.46	7.10	6.42	7.10
IDS, F_1, F_2	5.35	5.93	5.56	6.07
ADS, F_1, F_2	5.99	5.00	5.59	6.72
pairwise F-score, by register	A > I	I > A	A > I	A > I
by dimensions	$\forall$ > F	$\forall$ > F	$\forall$ > F	$\forall$ > F
classification accuracy, IDS	.639	.694	.708	.707
ADS	.732	.664	.714	.773
	A > I	I > A	A > I	A > I

Table 33: A summary of the performance of the clustering models and logistic regressions. This table indicates the performance of clustering models for IDS and ADS data, as well as formant (F) and multidimensional ($\forall$) data, through the number of clusters, K, and pairwise F-scores. Register-specific differences in the performance of the logistic regressions are indicated by measures of classification accuracy.

Both of these analyses captured the absolute discriminability of vowel distinctions in the multidimensional acoustic data that was sampled from the naturalistic corpus of IDS and ADS. Both of these models indicated that the distributional properties of the input were ambiguous and that the input did not allow for learners to identify the number of categories in the vowel system or their parameters. This poor performance was observed even though this analysis considered multidimensional acoustic data sampled from IDS and ADS: this stood in contrast to previous computational models that have typically considered formant distributions sampled from one of these two registers. Register-specific effects associated with hyperarticulation in IDS did not improve the performance of these models. This result was unsurprising given that the acoustic analysis in chapters 3 and 4 indicated a lack of enhancement in IDS for three of the four speakers in the corpus. The use of multidimensional acoustic data resulted in improved performance in the models that were applied to data from each register. Clustering models had higher pairwise F-scores when they were applied to the multidimensional data in comparison to low dimensional data. Similarly, logistic regressions had significant coefficients for vowel duration and patterns of spectral change for the majority of vowels, indicating that these additional dimensions were predictors of vowel quality. These results therefore supported the claim that formant analyses may overstate the ambiguity of the input (Swingley, 2009). Despite this facilitative effect, the multidimensional input still only provided ambiguous information about the vowel system of American English.

Though the results of the clustering models and logistic regressions were parallel to one another, differences in the assumptions of these approaches highlighted cases where the discriminability of vowel categories was affected by their frequency in the input. In the clustering models, the frequency of categories mirrored their frequency in the input while the frequency of each category was balanced in the regression analyses.

An analysis of Japanese IDS has indicated that low frequency categories may not be detected through the use of distributional learning (Bion et al., 2013). This study highlighted that although vowel duration was a strong predictor of phonological vowel length in Japanese, the distribution of this dimension only had a single mode. Long vowels were not represented by a separate mode since they formed 4% of the input. This effect of frequency indicates a case where measures of overlap do not sufficiently describe the distributional properties of the input (Cristia and Seidl, 2014). Comparisons of the results of the clustering models and logistic regressions highlighted further examples of this type. The logistic regressions had high F-scores for /ʊ/ and /ɔɪ/, indicating that the identity of these categories were predictable from their acoustic properties. Despite this, the clustering models never identified individual clusters that corresponded to these categories. Frequency may have negatively affected the discriminability of these categories since /ʊ/ was less frequency than /u/ and /oʊ/ while /ɔɪ/ was the least frequent vowel in the corpus. These effects suggested that frequency must be considered when assessed the discriminability of categories in the input. Measures of overlap and the outcomes of regression analyses may have overstated the discriminability of these vowels.

Though these results indicated that distributional learning was not a viable learning mechanism in infancy, they should be interpreted with care. These results indicate that the set of acoustic dimensions that were considered in the acoustic analyses that were presented were not sufficient to discover vowel categories which corresponded to the vowel phonemes of American English. The current analysis does not rule out the possibility that distributional learning may viable if an even broader set of acoustic dimensions were to be considered. Similarly, it does not indicate the extent to which this mechanism can identify acoustic categories that do not correspond to phonemes. The fact that distributional learning over the acoustic input is insufficient to explain perceptual attunement should not be interpreted as evidence that this mechanism cannot play a useful role in theories of perceptual development. To state that distributional learning is the sole explanatory mechanism behind perceptual attunement would be to misrepresent these theories. Approaches such as NLM (Kuhl et al., 2008) and PRIMIR (Werker and Curtin, 2005) integrate distributional learning into their theories alongside infants' emergent generalisations about the lexical items and phonotactic properties of their native language. Computational models have shown accurate categorisation when the simultaneous acquisition of multiple levels of structure is implemented alongside phonetic category learning (lexical items: Feldman et al., 2013; allophonic rules: Dillon, Dunbar, and Idsardi, 2013; semantic contexts: Frank, Feldman, and Goldwater, 2014). These models demonstrate that acquiring multiple levels of linguistic structure in parallel may facilitate perceptual attunement. Although the noisy data in the input may only allow for a partial solution to each separate domain, each level can bootstrap the processing and recognition of others and thus ensure successful acquisition.

6 General discussion

In this thesis, I have presented an acoustic analysis and a series of computational modules which address two research questions regarding the development of speech perception. The first research question considered the extent to which the acoustic properties of IDS are consistent with the hyperarticulation hypothesis (Bernstein Ratner, 1984; Kuhl et al., 1997). This hypothesis claims that the properties of IDS vowel production facilitate perceptual attunement in infancy and forms part of a larger functionalist explanation of this register's properties. The comparative acoustic analyses in chapters 3 and 4 indicated that there was a lack of enhancement in IDS, challenging current assumptions regarding the relative discriminability of vowels across registers. Though dispersion was greater in IDS than ADS, caregivers' vowel productions were more variable and overlapped to a greater extent in this register relative to ADS. These results demonstrated a need to report variance-sensitive measures in this domain and prompted a reconsideration of the functionalist explanation of IDS vowel production. The second research question considered the extent to which distributional learning can explain perceptual attunement in infancy (Maye, Werker, and Gerken, 2002). This mechanism has been attributed a central role in perceptual development as experimental tasks have indicated that the statistical regularities of the input modulate perceptual behaviour in infancy. The statistical models that I applied to samples of IDS and ADS in chapter 5 challenged this view as they demonstrated that categories which correspond to the vowel phonemes of American English cannot be identified from the distributional properties of the acoustic input. These analyses indicated that current theories of perceptual development must be reformed by assessing the mechanisms behind this process, the units being learnt, and the aspects of the input which the learner attends to.

6.1 Thesis summary

Chapter 1 located these two interrelated research questions and highlighted their relation to one another. Investigations of the hyperarticulation hypothesis closely capture the statistical regularities of the input and therefore can be used to determine the extent to which it enables distributional learning in infancy. Models which replicate learning in infancy highlight the learning outcomes that can be associated with the alterations that caregivers make to their speech in IDS and thus indicate whether these features facilitate perceptual attunement. The section established the findings and limitations in previous work in order to illustrate the goals of the thesis. With regard to the first research question, previous acoustic studies have not always supported the hyperarticulation hypothesis and measures of discriminability do not fully capture the distributional properties of the input. With regard to the second research question, the poor performance of computational models which replicate infant learning have suggested that distributional learning may not trivially extend from laboratory contexts to infants' day-to-day linguistic experience.

Chapter 2 reviewed the current literature concerning perceptual attunement, the acoustic properties of IDS, and distributional learning in infancy. This discussion first established how the perception of speech sounds changes within the first year of life, locating both the properties of the input and the use of distributional learning in a broader theoretical context. I then summarised the experimental evidence concerning the availability of distributional learning in infancy in order to highlight which properties of the input affect speech perception. I contrasted successful learning in experimental settings with the poorer performance of computational models that replicate this learning process in order to highlight that the linguistic input may not support the use of this mechanism. I further argued that models which have successfully identified categories have either depended on assumptions which simplify the learning task or the use of inferences from other levels of linguistic structure. In an overview of the hyperarticulation hypothesis, I outlined comparative acoustic analyses of IDS and ADS and the methods that have been used to capture the relative discriminability of these registers. This discussion indicated that studies have not unanimously supported the claim that vowel categories in IDS are more discriminable than their ADS counterparts. I further argued that conflicting findings can be attributed to the fact that that the most commonly adopted measure, the area of the vowel space, does not sufficiently describe the distributional properties of the input. I therefore advocated for the use of variance-sensitive measures in this domain and further highlighted that studies which have adopted these measures have not provided evidence of contrast enhancement in IDS.

F_1, F_2	ALI	ANN	CIN	GAI
area	I > A	I > A	I > A	I > A
peripherality	ns	ns	I > A	ns
2D dispersion	I > A	I > A	I > A	I > A
F_1 dispersion	ns	ns	I > A	ns
F_2 dispersion	ns	I > A	I > A	I > A
F_1 variance	I > A	ns	I > A	I > A
F_2 variance	I > A	ns	I > A	I > A
D(a)	A > I	ns	ns	A > I
F_1 D(a)	A > I	A > I	A > I	A > I
F_2 D(a)	A > I	I > A	I > A	A > I
S2	ns	ns	I > A	I > A
S2 ratio	ns	ns	ns	ns
mean F_1	ns	A > I	ns	A > I
mean F_2	ns	ns	ns	ns

Table 34: A summary of the acoustic measures of discriminability (and positive affect) for the analysis of F_1 and F_2. This table indicates the presence and direction of any significant effects which were identified through the use of Wilcoxon signed-rank tests.

Chapter 3 presented a comparative acoustic analysis of IDS and ADS which operationalised vowel quality through measures of the first two formants. This analysis considered a speech corpus that was originally collected in Bernstein Ratner (1984) which details four caregivers in interactions with their infants and with an adult experimenter. This analysis considered over a thousand vowel tokens from speaker's productions of both IDS and ADS. This formant analysis described register-specific differences in the central tendency and variance of the fifteen American English vowel categories. Table 34 indicates presence and direction of any register-specific differences for the measures of discriminability (and maternal affect) that were presented in this chapter. This table was originally presented as table 19 in 3.5. While vowel space expansion and dispersion were consistent with enhancement in IDS, variance-sensitive measures did not support the hyperarticulation hypothesis. The within-category variance of F_1 and F_2 was greater in IDS than ADS. Because of this, IDS vowels had a comparable or greater degree of overlap than those in ADS. Given that this analysis considered a broader set of categories and measures of discriminability than previous studies, this chapter provided a more exhaustive account of the distributional properties of these two registers. A lack of enhancement in IDS was consistent previous variance-sensitive measures of discriminability that have been reported in this domain.

Chapter 4 extended this comparative acoustic analysis of IDS and ADS by assessing register-specific differences in the central tendency and variance of F_3, vowel duration, and patterns of spectral change. The use of multidimensional data addressed claims that these additional acoustic dimensions provide stronger evidence of hyperarticulation than formant analyses (Eaves Jr. et al., 2016) and/or mitigate the ambiguity that has been seen in formant distributions (Swingley, 2009). Tables 35–38 summarise register-specific differences in the measures of discriminability (as well as maternal affect and speech rate) that were presented in this chapter. These tables were originally presented as tables 20–23 in 4.6. As with F_1 and F_2, measures of dispersion were generally greater in IDS than ADS. However, the variance of IDS vowels was again greater than ADS vowels for these additional dimensions. Measures of the degree of overlap therefore only indicated an effect of enhancement in IDS for one of the four speakers. Conversely, effects of deterioration were observed in the IDS productions of two speakers. These results aligned with the previous chapter as the lack of enhancement in IDS provided evidence against the hyperarticulation hypothesis. This lack of enhancement also provided evidence against the claim that consideration of multidimensional acoustic data provides stronger evidence of enhancement in IDS than formant analyses do. This observation validated the previous results of formant analyses which have not provided evidence of enhancement, indicating that such results cannot be dismissed as false negatives with regard to the hyperarticulation hypothesis.

Chapter 5 presented a series of clustering models and logistic regressions which assessed the extent to which distributional learning can explain perceptual attunement in infancy. Table 39 summarises how register-specific differences in vowel quality and

F_3	ALI	ANN	CIN	GAI
dispersion	A > I	ns	I > A	I > A
variance	A > I	ns	I > A	I > A
D(a)	ns	I > A	ns	I > A
mean F_3	ns	I > A	I > A	ns

Table 35: A summary of the acoustic measures of discriminability (and positive affect) that were applied to the third formant. This table indicates the presence and direction of any significant effects which were identified through the use of Wilcoxon signed-rank tests.

spectral change	ALI	ANN	CIN	GAI
dispersion	ns	I > A	I > A	I > A
ΔF_1 dispersion	ns	I > A	I > A	I > A
ΔF_2 dispersion	ns	I > A	I > A	I > A
ΔF_1 variance	I > A	ns	I > A	I > A
ΔF_2 variance	I > A	ns	I > A	I > A
D(a)	A > I	I > A	I > A	A > I
ΔF_1 D(a)	A > I	I > A	ns	A > I
ΔF_2 D(a)	A > I	I > A	I > A	A > I

Table 36: A summary of the measures of discriminability that were applied to patterns of spectral change.

log vowel duration	ALI	ANN	CIN	GAI
dispersion	I > A	I > A	I > A	I > A
variance	I > A	I > A	I > A	I > A
D(a)	ns	I > A	ns	ns
mean log duration	I > A	I > A	I > A	I > A

Table 37: A summary of the measures of discriminability (and speech rate) that were applied to log vowel duration.

all dimensions	ALI	ANN	CIN	GAI
dispersion	ns	I > A	I > A	I > A
D(a)	A > I	I > A	ns	A > I

Table 38: A summary of the measures of discriminability that were applied to the multidimensional space that was defined by the first three formants, patterns of spectral change, and log vowel duration.

	ALI	ANN	CIN	GAI
mean K, IDS, all dimensions	5.63	7.76	7.60	8.13
ADS, all dimensions	6.46	7.10	6.42	7.10
IDS, F_1, F_2	5.35	5.93	5.56	6.07
ADS, F_1, F_2	5.99	5.00	5.59	6.72
pairwise F-score, by register	A > I	I > A	A > I	A > I
by dimensions	$\forall$ > F	$\forall$ > F	$\forall$ > F	$\forall$ > F
classification accuracy, IDS	.639	.694	.708	.707
ADS	.732	.664	.714	.773
	A > I	I > A	A > I	A > I

Table 39: A summary of the performance of the clustering models and logistic regressions. This table indicates the performance of clustering models for IDS and ADS data, as well as formant (F) and multidimensional ($\forall$) data, through the number of clusters, K, and pairwise F-scores. Register-specific differences in the performance of the logistic regressions are indicated by measures of classification accuracy.

the use of multidimensional data affected model performance. This table was originally presented as table 33 in 5.5. Clustering models replicated distributional learning in infancy by assigning vowel tokens to an unknown number of categories on the basis of similarity. By contrast, multinomial logistic regressions indicated the inferences that an ideal observer could draw from these samples of acoustic data. Each of these models indicated that the statistical properties of the input were not sufficient to identify native language vowel categories, limiting the viability of distributional learning as an explanatory mechanism in infancy. The clustering models typically recovered between six and eight categories rather than the fifteen vowel phonemes of American English. The logistic regressions confirmed this result as the classification accuracy of these models was not at ceiling for any speaker or register. The acoustic data therefore provided learners with ambiguous predictors of category identity. The models that were applied to multidimensional data outperformed those that were applied to formant distributions and thus supported the claim that multidimensional data can mitigate the ambiguity of formant data (Swingley, 2009). The observation of significant coefficients for dimensions other than F_1 and F_2 also confirmed that multidimensional data supported the identification of vowel categories. These models also confirmed that the acoustic properties of IDS and ADS that were reported in chapters 4 and 5 had an impact on the viability of distributional learning. Register-specific differences in the performance of the clustering models and the logistic regressions aligned with the measures of overlap which were reported in chapters 4 and 5 rather than indicating improved performance in IDS. These results supported the use of D(a) in future comparative acoustic analyses of IDS and ADS. The regressions also indicated the discriminability of individual categories, demonstrating that peripheral vowels were more discriminable than central vowels and that more frequent vowels were more discriminable than less frequent ones.

6.2 The hyperarticulation hypothesis

The hyperarticulation hypothesis forms part of a larger functionalist claim which maintains that IDS bears the similar properties across the world's languages as caregivers adapt their speech in order to promote language learning. This hypothesis specifically claims that the distributional properties of IDS vowels enable perceptual attunement in infancy. Comparative acoustic studies in this domain have not unanimously provided evidence in support of this claim and I have presented new empirical data which clarifies these results. Additionally, I have forwarded new methodologies which can be used to assess the relative discriminability of IDS and ADS. On the basis of these acoustic analyses, I have argued that mixed results in this domain can be attributed to the fact many studies have applied measures of discriminability which do not fully capture the distributional properties of the input to subsets of the relevant of categories and acoustic dimensions that are present in the input. I have argued that the area of the vowel space is not informative of the properties of the system as a whole, even though it is the most commonly adopted measure in this domain. Though the effects of vowel space expansion that were originally observed in American English, Russian, Swedish IDS (Kuhl et al., 1997) have been replicated in other languages (American English: Cristia and Seidl, 2014; Hartman, Bernstein Ratner, and Newman, 2016; Australian English: Burnham, Kitamura, and Vollmer-Conna, 2002; Kalashnikova, Carignan, and Burnham, 2017; Xu et al., 2013; British English: Uther, Knoll, and Burnham, 2007; Japanese: Miyazawa et al., 2017; Mandarin: Liu, Kuhl, and Tsao, 2003), a further set of studies have observed that the area of the IDS vowel space may be either comparable to or smaller than that of the ADS vowel space (American English: Burnham et al., 2015; Cantonese: Xu Rattanasone, Burnham, and Reilly, 2013; Danish: Bohn, 2013; Dutch: Benders, 2013; Norwegian: Englund and Behne, 2006; French, British English, Japanese: Dodane and Al-Tamimi, 2007). In addition to this, I have forwarded three separate limitations that reveal that the area of the vowel space provides a poor indicator of how the distributional properties of the vowel system as a whole differ across registers. Firstly, this measure only indicates the central tendency of the three point vowels. In order to determine whether the effects of expansion occur across the system as a whole, it is necessary to compare measures of peripherality and dispersion for all vowels in the system across registers. Secondly, this measure does not detect register-specific differences in the within-category variance of these categories. Variance-sensitive measures are required in order to determine whether the acoustic input presents learners with the individual modes for each category which enable distributional learning (Maye, Werker, and Gerken, 2002). As highlighted in Cristia and Seidl (2014), greater dispersion in IDS only results in a lesser degree of overlap if within-category variance is comparable across IDS and ADS. Such assumptions are problematic as many comparative studies have found that IDS vowel production is more variable than ADS vowel production (American English, Russian and Swedish:

Kuhl et al., 1997; American English: Cristia and Seidl, 2014, Kirchhoff and Schimmel, 2005, McMurray et al., 2013; Dutch: Benders, 2013; Japanese: Miyazawa et al., 2017). Thirdly, comparative studies in this domain have been criticised for solely considering measures of the first two formants (Eaves Jr. et al., 2016). This critique proposes that caregivers may optimise native-language distinctions in higher dimensional space that the two-dimensional space that is described by F_1 and F_2. If this claim holds, multidimensional analyses of vowel production are needed in order to identify effects of enhancement in IDS.

In order to address each of these concerns, I will consider and evaluate the measures of discriminability that were presented in the comparative acoustic analyses of IDS and ADS from chapters 3 and 4. This analysis replicated previous studies by demonstrating an effect of vowel space expansion in IDS. The positive effect that this had on the dispersion of the three point vowels extended to the full set of fifteen American English vowels in IDS. However, I will argue that measures of the area of the vowel space are a poor indicator of the distributional properties of vowel categories and, by extension, the intentions of caregivers. The main flaw of this statistic is its inability to detect register-specific differences in within-category variance. I will argue that researchers in this domain should view the greater variability which I observed in vowels in IDS relative to ADS as a definitive property of vowel production in this register. This effect of variance provided evidence against the hyperarticulation hypothesis as IDS vowels did not show a lesser degree of overlap than their ADS counterparts, despite the observation of greater dispersion in this register. Variance-sensitive measures of discriminability are required in order to fully interpret the discriminability of categories in each register. This aligns with the results of previous studies as variance-sensitive measures have not demonstrated an effect of enhancement in IDS (Cristia and Seidl, 2014; Miyazawa et al., 2017). As multidimensional acoustic analyses provided only limited evidence of enhancement in IDS, I will argue that such methods are not required in order to interpret the register-specific differences in discriminability. Previous formant analyses which have failed to demonstrate an effect of enhancement therefore cannot be dismissed as false positives.

Measures of discriminability In order to facilitate a discussion of the measures of discriminability that have been adopted in this domain, I will restate how each of these statistics indicate an effect of enhancement in IDS. Greater values for the area of the vowel space, peripherality, and dispersion in IDS relative to ADS provide evidence of enhancement in speech addressed to infants. Conversely, greater within-category variance in IDS in comparison to ADS would provide evidence of contrast deterioration. Greater values for D(a) in IDS than ADS would also indicate an effect of enhancement in this register by indicating a lesser degree of overlap. Chapter 3 further reported measures of S2 (and the ratio of observed to maximal S2) which captured the relative orientation of paired categories. Greater values for these statistics in IDS than

ADS would indicate that speakers enhanced distinctions in this register by ensuring that categories differed in orientation from one another (Eaves Jr. et al., 2016). The ratio of observed to maximal S2 was additionally reported as greater values for S2 in IDS may have just reflected the greater variability of categories in this register.

I applied each of these measures to the fifteen vowels of American English and to the set of 105 distinctions that were formed by pairing these categories. I used statistical tests to diagnose register-specific differences in these measures for all of the categories or distinctions in the system. In addition to detecting effects of enhancement in IDS, these tests could also reveal two effects which are consistent with a lack of enhancement. Null results either indicate that the vowels from each register did not differ with regard to the relevant statistic or that a majority of categories did not exhibit the same effect across registers. Negative results further indicate that the discriminability of IDS vowels is poorer than those in ADS. These analyses therefore demonstrated whether register-specific differences in discriminability could be generalised across the system as a whole. Analyses which consider too few vowels have potentially provided incorrect conclusions in this regard. For example, the principle of maximal dispersion states it is necessary to consider all of the categories in an inventory in order to identify the arrangement in which maximises the inter-category distances across all vowels (Liljencrants and Lindblom, 1972; Lindblom, 1986). Excluding specific categories reduces the degree of overlap in the system by eliminating confusable alternatives for each vowel. Though the current analysis considered a broad range of categories, it considered a comparatively small population of caregivers. These comparative analyses should ideally reveal the same register-specific patterns across all four speakers. Such effects will be interpreted strongly while those associated with individuals will be interpreted with caution. The small sample size means that differences in each speaker's patterns of enhancement could not further be associated with factors such as the age or developmental level of the infant addressee.

Measures of the central tendency The measures of expansion, peripherality, and dispersion that were presented in chapters 3 and 4 enables a reinterpretation of previous studies which have observed vowel space expansion in IDS. This analysis provided evidence that was consistent with the hyperarticulation hypothesis as IDS distinctions exhibited greater dispersion than their ADS counterparts across the system as a whole. As the area of the vowel space was numerically larger in IDS than ADS for all four speakers, the current study replicates previous studies which have observed this effect in American English IDS and suggests that this effect is robust (Cristia and Seidl, 2014; Hartman, Bernstein Ratner, and Newman, 2016; Kuhl et al., 1997). This analysis indicated that inter-category Euclidean distances were greater in IDS than ADS for all four speakers. Measures of peripherality did provide evidence of a similar facilitative effect as only speaker CIN exhibited greater peripherality in IDS than ADS. This measure did not differ across registers for speakers ALI, ANN, or GAI. The current set of results provided further insights into previous studies which have presented measures of peripherality and dispersion as evidence against the hyperarticulation hypothesis. An analysis of American English IDS and ADS found that the degree of peripherality did not differ registers for a set of nine vowels (McMurray et al., 2013). The authors interpreted this effect as evidence that caregivers did not enhance vowel distinctions in IDS or, at least, that they prioritised the promotion of other aspects of the linguistic input over the enhancement of vowel distinctions. They further stated apparent effects of hyperarticulation are better explained as an epiphenomenon which arises from the prosodic differences that exist between IDS and ADS. The current set of results provided evidence against this interpretation of such effects of peripheralisation as it indicated that greater dispersion could occur in IDS while the peripherality of categories remained comparable across registers. The current analysis also stands in contrast to studies where measures of dispersion have indicated a lack of enhancement in IDS. Inter-category Euclidean distances were comparable across registers for two American English tense-lax distinctions (/i, ɪ/; /eɪ, ɛ/: Cristia and Seidl, 2014) and for distinctions between point vowels and two further vowel pairs in Danish (/i, e/; /oː, ɔː/: Bohn, 2013). This analysis of Danish found that the dispersion between /eː/ and /ɛː/ was lower in IDS than ADS. In contrast to the current analysis, these previous analyses only indicated a lack of enhancement for a small set of distinctions. Such results can be interpreted as only being informative for those specific distinctions rather than the system as a whole. Such cases illustrate the need to consider the discriminability of all of the categories in a system in order to refute the hyperarticulation hypothesis. Cristia and Seidl (2014) acknowledged this interpretation and stated that their results only provide evidence against a variant of the hyperarticulation hypothesis which states that caregivers enhance all distinctions in IDS. Disproving such a hypothesis is trivial, however. Even in the current set of results which did exhibit evidence of greater dispersion in IDS than ADS, this effect was not observed for all 105 distinctions.

Comparisons of the central tendency across registers did not support the claim that IDS vowel production may be better explained as expressing positive affect rather enhancing vowel distinctions (Benders, 2013). This account predicts that the value for each formant is greater in IDS than ADS across all vowels in the system. High frequencies indicate positive affect as they imply a small body size and a low threat level, following the frequency-size relationship (Ohala, 1980, 1984). The current analysis found no consistent effect of formant raising in IDS. Though the third formant was greater in IDS than ADS for speakers ANN and CIN, the same effect was not observed for other speakers or formants. The value of F_1 did not differ across registers for speakers ALI and CIN while speakers ANN and GAI had lower values for this formant in IDS than ADS. The value of F_2 did not differ across registers for any of the four speakers. Given that patterns of formant raising were less consistent than the register-specific differences in dispersion, I conclude that maternal affect does not provide a more apt account of IDS vowel production than the hyperarticulation hypothesis does.

Variance-sensitive measures Though measures of inter-category Euclidean distances were consistent effect of enhancement in IDS, variance-sensitive measures captured the distributional properties of IDS and ADS and prompted a reinterpretation of the relative discriminability of the two registers. Measures of within-category variance were greater in IDS than ADS to the extent that IDS vowels did not show a lesser degree of overlap than ADS vowels. These measures provided evidence against the hyperarticulation hypothesis. This effect was robust as within-category variance was greater in IDS than ADS across multiple speakers and acoustic dimensions. Though this finding indicated that speakers did not enhance distinctions in IDS, it is unsurprising since multiple studies have reported that the within-category variance of the first two formants is greater in IDS than ADS (American English, Russian and Swedish: Kuhl et al., 1997; American English: Cristia and Seidl, 2014, Kirchhoff and Schimmel, 2005, McMurray et al., 2013; Dutch: Benders, 2013; Japanese: Miyazawa et al., 2017). In my view, heightened variance should be considered as a definitive feature of IDS vowel production and incorporated into all comparative analyses in this domain. Measures of D(a) indicated that there was either a lack of enhancement or an active effect of deterioration in IDS. Register-specific differences in variance therefore had a greater larger impact on the discriminability of vowel categories than register-specific differences in their central tendency. As highlighted in Cristia and Seidl (2014), the observation of greater dispersion only indicates a lesser degree of overlap if it can be assumed that within-category variance remains constant across register. As the current analysis indicated that variance actually has a negative impact of measures of overlap, I forward the possibility that studies which have reported measures of the central tendency of vowel categories have incorrect conclusions regarding the hyperarticulation hypothesis. The current lack of enhancement in IDS aligns closely with previous studies which have reported measures of overlap. The value of D(a) for a pair of American English distinctions was not greater in IDS than ADS (/i, ɪ/; /eɪ, ɛ/: Cristia and Seidl, 2014) while an analysis of ten distinctions between the five vowel phonemes of Japanese found that the Mahalanobis distance between these distinctions in IDS was less than in ADS (/i, ɛ, ɑ, o, ɯ/: Miyazawa et al., 2017). This study of Japanese provided further evidence for a lack of enhancement in IDS as the degree of overlap in IDS and ADS was greater than that which was observed in clear speech.

Comparisons of the relative orientation of categories across registers further supported the interpretation that register-specific differences in variance had a negative impact on the discriminability of IDS distinctions. These measures provided evidence against an alternative interpretation of the data where caregivers enhance distinctions between high variance categories by orienting them orthogonally to each other in acoustic space (Eaves Jr. et al., 2016). Measures of S2, which indicated the relative orientation of categories, were greater in IDS than ADS for speakers CIN and GAI but did not differ across registers for speakers ALI or ANN. Though greater values indicated greater differences in orientation, further analyses indicated that this interpretation was spurious. The ratio of observed S2 to maximal S2 did not differ across registers for any of the four speakers. These results indicated that greater values of S2 in IDS only reflected the greater variance of vowels in this register. Given that speakers did not modulate the relative orientation of IDS vowel categories, the heightened variance of IDS vowels must be interpreted as evidence against the hyperarticulation hypothesis.

Multidimensional acoustic data The multidimensional acoustic analysis of IDS and ADS in chapter 4 provided new empirical information concerning register-specific differences in caregivers' realisation of the third formant, vowel duration and patterns of spectral change. This chapter observed register-specific differences in the same set of measures which generally aligned with the results of the formant analysis. Though dispersion was greater in IDS than ADS across many acoustic dimensions, within-category variance was also greater in this register than in ADS. Measures of the degree of overlap therefore indicated a lack of enhancement in IDS. This analysis informed two proposals which have advocated for the use of multidimensional acoustic data when analysing the discriminability of IDS vowels. The first states that multidimensional acoustic data provides stronger evidence of enhancement in IDS than formant distributions (Eaves Jr. et al., 2016). The current analysis provided evidence against this claim as the results of the formant analyses and multidimensional analyses were generally comparable. Formant analyses which have not indicated an effect of enhancement in IDS therefore cannot be reinterpreted as false negatives with regard to the hyperarticulation hypothesis. The second states that multidimensional acoustic data provides more information about vowel distinctions and therefore mitigates the cases of overlap which have been observed in formant analyses (Swingley, 2009). As this proposal concerns the absolute discriminability of categories in a single register, the comparative acoustic analysis of IDS and ADS did not address it directly. This proposal will be evaluated in a later discussion of the statistical models from chapter 5 and their implications for the viability of distributional learning.

The results of multidimensional acoustic analysis of IDS and ADS revealed effects of dispersion, variance, and overlap that were comparable to those which were observed in two-dimensional formant space. Measures of dispersion indicated that caregivers made alterations to the central tendency of IDS vowel categories that were generally consistent with the hyperarticulation hypothesis. Inter-category Euclidean distances were generally greater in IDS than ADS across these additional acoustic dimensions. Speakers CIN and GAI had greater dispersion for F_3 in IDS than ADS: conversely, speaker ALI had greater dispersion in ADS than IDS for this dimension. Speakers ALI, ANN, and GAI had greater dispersion in IDS than ADS for patterns of spectral change while all four speakers showed this facilitative effect for log vowel duration. This analysis further indicated that within-category variance was greater in IDS than ADS for these additional dimensions. Speakers ALI, CIN, and GAI had greater within-category variance in IDS than ADS for measures of the third formant and both measures of spectral change. Vowel duration had greater within-category variance in IDS than ADS for all four speakers. The inverse effect, where variance was greater in ADS than IDS, was only observed in a single case, namely for F_3 in speaker ALI's data. As in the previous chapter, measures of D(a) indicated that within-category variance had a greater impact on the discriminability of categories across registers than dispersion did. Measures of D(a) indicated a lack of enhancement for three of the four speakers as this statistic did not differ across registers for CIN while both ALI and GAI showed greater overlap in IDS than ADS. The greater values of D(a) in IDS than ADS for ANN was the only effect of enhancement that was observed in IDS and this effect followed from the observation of greater dispersion in IDS and comparable within-category variance across registers.

Given that the analyses in chapters 3 and 4 identified similar register-specific effects, this multidimensional analysis did not support the claim that multidimensional data provides stronger evidence of enhancement in IDS than formant analyses do Eaves Jr. et al. (2016). Both the formant and multidimensional analysis indicate a lack of enhancement in IDS for speakers ALI, CIN, and GAI. Speaker ANN is the sole case where a multidimensional analysis revealed an effect of enhancement in IDS that was not detected by the formant analysis. The current set of results align with the those of previous comparative acoustic analyses of IDS and ADS which observed a lack of enhancement in IDS in spite of the use of multidimensional acoustic data (Cristia and Seidl, 2014; Martin et al., 2015). A comparative analysis of American English IDS and ADS assessed the discriminability of a pair of tense-lax distinctions through measures of the first two formants, vowel duration and patterns of spectral change (/i, ɪ/; /ɛ, eɪ/: Cristia and Seidl, 2014). Despite this multidimensional information, inter-category Euclidean distances and D(a) were not consistent with an effect of enhancement. A comparative analysis of all of the segmental distinctions in Japanese used a variant of cepstral coefficients to capture the entire spectral envelop (Martin et al., 2015). This analysis of Japanese revealed that there was a small but significant effect of contrast deterioration in IDS relative to ADS. Taken together, these previous results and those of the current analysis did not support the claim that multidimensional acoustic data provides stronger evidence of contrast enhancement than formant distributions do. Though the current analysis indicates that formant analyses do not capture all of the alterations that speakers make to vowel production in IDS, these analyses cannot be described as failing to observe effects of contrast enhancement. Multidimensional data may assist in the recognition of categories from the input but it does not reveal additional effects of enhancement. Because of this, formant analyses which provide evidence against the hyperarticulation hypothesis must be interpreted as legitimate characterisations of this register rather than being dismissed as false negatives.

Insights from computational models The current consideration of the hyperarticulation hypothesis was novel as the discriminability of vowels across registers was assessed through both acoustic analyses and computational models. The acoustic analysis identified the existence and direction of any register-specific differences in discriminability while the models indicated how these effects impacted the infant learner. To my knowledge, only one previous study has considered these two methods in parallel by pairing measures of peripherality with a logistic regression (McMurray et al., 2013). Specifically, this analysis of American English used regressions to indicate that a lack of global peripheralisation in IDS relative to ADS was consistent with a lack of enhancement. In the current set of models, differences in model performance across registers aligned with the values of D(a) that were observed for each register in multidimensional acoustic space. The clustering models had lower pairwise F-scores in IDS than ADS for ALI, ANN, and GAI for whom there was not a lesser degree of overlap in IDS relative to ADS. Conversely, pairwise F-scores were greater in IDS than ADS for ANN who had a lesser degree of overlap in IDS than ADS. The classification accuracy of the logistic regressions indicated similar register-specific differences in model performance. Classification accuracy was lower in IDS than ADS for the three speakers who showed a lack of enhancement in IDS while this statistic was greater in IDS than ADS for ANN. Though differences in D(a) aligned well with the performance of these models across registers, similar relationships were not observed for measures of the area of the vowel space or dispersion. The current analysis therefore supported the use of D(a) as a measure of discriminability in this domain as this statistic could be directly associated with the performance of models which replicated distributional learning in infancy. Evidence of this type is great importance given that the measures of discriminability have been criticised for being overly dependent on researchers' intuitions of which properties of the input enable learning in infancy (Eaves Jr. et al., 2016). These computational models supported the intuition that input with a lesser degree of overlap enables the use of a statistical mechanism which depends on the identification of modes in a frequency distribution.

This analysis found that differences in the discriminability of vowel across registers had a comparatively small effect on the performance of these models. Clustering models typically recovered between six and eight categories regardless of the register that they were applied to and the classification accuracy of the logistic regressions never exceeded 80% for either register. The observation of a lesser degree of overlap in a certain register should not be interpreted as evidence that learners can trivially identify categories from that register. Similarly, registers with a greater degree of overlap should still be interpreted being a source for relevant statistic inferences regarding the vowel category distinctions of American English.

The interpretation of register-specific effects In summary, these comparative acoustic analyses of IDS and ADS present new empirical evidence which does not align with the hyperarticulation hypothesis. This analysis reveals how measures of discriminability which focus on the the central tendency of categories can provide a misleading view of the discriminability of these two registers. The current analysis replicates previous studies by identifying an effect of vowel space expansion in IDS. This can be generalised across the system as a whole as inter-category Euclidean distances tend to be greater in IDS than ADS for a set of 105 distinctions. This analysis does not provide evidence that peripheralisation differs across registers, however. Despite the existence of these facilitative effects, the primary contribution of this analysis is to highlight the importance of within-category variance in these comparative analyses. Variance-register measures indicate a lack of enhancement which prompts a reinterpretation of the previously discussed measures. The high within-category variance of vowels in IDS is a feature that has now been reported across large number of studies of IDS vowel production (American English, Russian and Swedish: Kuhl et al., 1997; American English: Cristia and Seidl, 2014, Kirchhoff and Schimmel, 2005, McMurray et al., 2013; Dutch: Benders, 2013; Japanese: Miyazawa et al., 2017). Despite this, comparatively few studies have observed how it can negatively affect the discriminability of vowel distinctions. Measures of the degree of category overlap generally indicate a lack of enhancement in IDS relative to ADS. I view this as a major problem in this domain and argue that a failure to account for this aspect of the input has contributed to the mixed results regarding the hyperarticulation hypothesis. If, as I have argued, measures of the central tendency of categories do not reliably indicate the discriminability of vowel categories, then previous studies which have presented mixed evidence through the use of these measures need to be reinterpreted. A more uniform pattern of results may emerge if the samples of IDS and ADS that were considered in these studies could be reanalysed with variance-sensitive measures.

Though register-specific differences in dispersion align with the predictions of the hyperarticulation hypothesis, the effects associated with variance and category overlap necessitate a reconsideration of this functionalist approach. One interpretation of the current set of data would be to view these results as sufficient to refute the hyperarticulation hypothesis. However, stating that caregivers do not enhance vowel distinctions in IDS would not predict the occurrence of a consistent effect of dispersion in this register or, conversely, cases of contrast deterioration which arise through greater within-category variance. If IDS is solely characterised as having a lack of enhancement, the statistical properties of IDS and ADS should be comparable to another other. In order to address this, I propose that the effects of dispersion and variance are better explained as having distinct causes. Specifically, the greater dispersion of IDS vowels is the result of caregivers' intentions to promote language learning. By contrast, the heightened variance of this vowel may be associated with other highly variable aspects of the input. IDS has been characterised as being variable in nature, dependent on the age and developmental level of the infant addressee (Soderstrom, 2007). This characterisations further note that in addition to bearing short utterances with simplified syntax, IDS also feature longer utterances which are apparently self-directed. The greater variance of the first two formants has further been associated with the observation that F_0 is more variable in this register than in ADS (Cristia and Seidl, 2014). Such factors must be investigated closely in IDS and ADS in order to determine the extent to which they explain the variance that has been observed in measures of the first two formants as well as other acoustic dimensions across registers.

Though greater within-category variance may hinder the use of distributional learning in infancy, the high variance of vowels in IDS has been stated to facilitate the emergence of robust vowel perception (Cristia and Seidl, 2014; Miyazawa et al., 2017). It is important to distinguish effects which ensure the emergence of robust category sensitivity from those which ensure that infants can establish the existence of distinctions in the input. During perceptual attunement, high variance input hinders the identification of individual modes within a frequency distribution and thus limits the viability of distributional learning. By contrast, positive effects of variable input can be observed in tasks where infants have to employ their knowledge of the set of distinctions that exist in their native language. Infants show sharper vowel discrimination after training on tokens with a pitch contour in comparison to training on tokens with level pitch (Trainor and Desjardins, 2002). Exposure to high variance experimental stimuli also facilitates the identification of object-form mappings. Infants associated the forms /puk/ and /buk/ with novel objects with greater success when they were trained on input from multiple speakers in comparison to input from a single speaker (Rost and McMurray, 2009). In a parallel fashion, infants who heard /t/ and /d/ across multiple phonological environments were able to map the forms /tɑ/ and /dɑ/ to novel objects more easily than infants who heard these sounds in a single environment (Thiessen, 2011). Though claiming that highly variable input can facilitate learning may seem paradoxical, this claim focuses on the high variability of units of linguistic structure that are independent from the acoustic dimensions which indicate native language distinctions. Naturally, increasing the variability of relevant acoustic dimensions will hinder perceptual attunement in infancy.

Given that this discussion has proposed a novel interpretation of register-specific differences in vowel quality across registers, it is important to closely consider factors which may have affected comparisons of the central tendency and variance of categories across IDS and ADS. The current analysis considered a large naturalistic corpus of both IDS and ADS which strongly resembles the type of day-to-day input that infants are exposed to. The acoustic analyses from chapters 3 and 4 considered over a thousand tokens from each combination of caregivers and registers. This speech was produced spontaneously and vowel tokens therefore occurred across multiple prosodic contexts, word types and phonological environments. This stands in contrast to many studies which have considered IDS and ADS which were produced in reading tasks Dodane and Al-Tamimi (2007) and McMurray et al. (2013) or that were elicited with a fixed set of toys (Kuhl et al., 1997). These samples of acoustic data have typically featured fewer than a hundred tokens which occur in a limited number of word types and environments. Because of this, register-specific differences in vowel quality must be interpreted with respect to a trade-off between an ability to transparently describe infants' typical linguistic experience and an ability to control for factors which may affect the quality of vowels independent to the identity of the addressee. I acknowledge that the high variance of vowel categories that I have reported here could be partially attributed to an inability to control for the word type, prosodic position, lexical stress that vowels occurred in. Vowels tokens in IDS could be more variable than those in ADS because caregivers engaged in a greater range of activities when interacting with infants than they did with the adult experimenter. Caregivers' IDS productions can also be expected to vary across sessions as a function of the age or developmental level of their infant unlike their ADS productions. The current analysis does not fully address the claim that apparent effects of hyperarticulation may be better explained as a side-effect of the differences in prosodic structure of IDS and ADS (McMurray et al., 2013). Previous studies have supported this claim, for example, by demonstrating that measures of VOT in American English did not differ across registers once prosodic factors have been controlled for (McMurray et al., 2013). A comparative analysis of vowels in American English IDS and ADS similarly indicated that the peripherality of categories was better predicted by stress and prosodic position than by register (Wang, Seidl, and Cristia, 2015).

In order to more fully assess the hyperarticulation hypothesis, future comparative acoustic analyses in this domain must observe and explain register-specific differences in the central tendency and variance of vowel categories across registers. In order to gaining further insights into the intentions behind caregivers' vowel productions, I advocate the consideration of samples of IDS that have been produced both spontaneously and under controlled conditions. Considering samples of spontaneous speech is necessary in order to understand the limits of variation that are apparent within each register. The results of controlled phonetic experiments, by contrast, would determine the extent to which any register-specific differences can be attributed solely to the addressee rather than any differences in the prosody, sentence structure or lexical items that are used across the two registers. Though controlling for such factors provides a more transparent view of vowel production across registers, these effects must also be considered from the perspective of the infant learner when assessing effects of enhancement. If multiple factors must be closely controlled for in order to identify a phonetic effect which can enable learning, then such an effect is only relevant if the infant learner is sensitive to and able to compensate for that same set of factors. It is therefore necessary to interpret any facilitative effects which can be observed in IDS against the additional linguistic knowledge that learner must have in order to exploit them.

6.3 Distributional learning in infancy

Experimental tasks which examine distributional learning in infancy have shown that exposure to the statistical regularities of the acoustic input can shape infants' perceptual behaviours within the first year of life (Maye, Werker, and Gerken, 2002). This mechanism has been attributed a central role in theories of perceptual development since it makes explicit predictions about how the properties of the input modulate infants' sensitivity to category distinctions. The viability of this mechanism outside of laboratory contexts depends on an assumption that the frequency distribution of the acoustic input presents learners with individual modes which correspond to each phonetic category of their native language. Computational models which process the acoustic input in this way have been forwarded as providing support for the use of this mechanism. However, it is important to closely consider the extent to which these resemble learning in infancy. Successful cases of learning are problematic as they simplified the infant learning task by specifying the number of categories to be learnt in advance (Kornai, 1998), by considering only a subset of the system (Benders, 2013; Moeng, 2016; Vallabha et al., 2007), or by making both of these assumptions (Adriaans and Swingley, 2012, 2017; de Boer and Kuhl, 2003; Kirchhoff and Schimmel, 2005). By contrast, models which identify both the number of categories in the system as well as their identity show much poorer performance, suggesting that this mechanism does not allow learners to trivially identify phonetic categories (Antetomaso et al., 2016; Feldman et al., 2013; Frank, Feldman, and Goldwater, 2014; Mooney, 2015). Models of this type may have drawn incorrect conclusions regarding the viability of distributional learning, however, as clustering models of this type have only ever been applied to formant distributions that were sampled from a single register. Data from a single register may be insufficient to characterise the use of distributional learning in infancy as register-specific differences in discriminability could improve the performance of these models. Similarly, the use of multidimensional acoustic data has been proposed as a method of mitigating the ambiguity that has been seen in formant distributions and therefore may reveal a similar improvement in performance (Swingley, 2009).

The clustering models and logistic regressions that were presented in chapter 5 addressed these concerns and provided an overview of the distributional properties of caregivers' speech. The clustering models failed to replicate learning in infancy and thus indicated that previous models which learn both the number and identity of categories provide a genuine indicator of the viability of distributional learning in infancy. Specifically, these models indicate that distributional learning is not viable as an explanatory mechanism with regard to perceptual attunement in infancy. The clustering models typically identified between six and eight clusters rather than the full set of fifteen phonemic categories. This analysis provided support for the claim that additional acoustic dimensions mitigate the cases of overlap as models that were applied to multidimensional data outperformed those that were applied to formant distributions (Swingley, 2009). It should be noted, however, that the consideration of a broader range of acoustic dimensions did not trivialise the learning task as these models still failed to identify all fifteen categories. This analysis also indicated that the consideration of IDS data does not enable the use of distributional learning in infancy. Comparisons of the IDS and ADS models indicated differences in model performance which aligned with the lack of enhancement that was observed in the acoustic analyses. As there was no consistent effect of hyperarticulation in IDS, the properties of this register did not support the use of statistical learning mechanisms. Logistic regressions furthered these findings and indicated that the acoustic properties of vowel production from either register provided learners with ambiguous predictors of vowel identity in American English. Though the regressions indicated that acoustic dimensions beyond the first two formants were significant predictors of category identity, the performance of these models indicated that the identity of categories could not be reliably predicted through their acoustic properties even with explicit training. The comparatively poor performance of these models necessitates a reconsideration of current theories and models of perceptual development and suggest that this learning task must incorporate mechanisms that support distributional learning in infancy. Current models of learning in which phonetic categories are learnt alongside lexical items (Feldman et al., 2013), semantic contexts (Frank, Feldman, and Goldwater, 2014), and phonological rules (Dillon, Dunbar, and Idsardi, 2013) shown improved performance and demonstrate how bootstrapping from other levels of linguistic structure may ensure accurate vowel categorisation in infancy.

Clustering models I will first consider the results of the current set of EM-based clustering models and discuss how these models indicated that distributional learning has limited viability as a explanatory mechanism for perceptual attunement in infancy. As these learning models typically identified between six and eight categories, they demonstrated that this mechanism did not enable learner to identify the fifteen phonemic vowels of American English. These results indicated that learning models which have made simplifying assumptions about the learning task have misrepresented the viability of this mechanism. Specifically, I will contrast the current set of models with previous approaches in order to highlight two simplifying assumptions which have enabled phonetic categories to be successfully identified. On the one hand, specifying the number of categories in the system provides the model with knowledge that the infant learner does not have. On the other, models which focus on a subset of categories rather than the inventory as a whole eliminate potential cases of overlap in acoustic space. By identifying how these assumptions misrepresent the learning task, the current results promoted the use of clustering models which have been applied to entire inventories and that do not specify the number of categories in the system (Antetomaso et al., 2016; Feldman et al., 2013; Frank, Feldman, and Goldwater, 2014; Mooney, 2015). I will argue that the poor performance of these models should be viewed as a genuine indicator of the viability of distributional learning in infancy. As the current set of models considered multidimensional data that was sampled from both IDS and ADS, the current analysis extended the findings of previous approaches which have solely considered formant data that was sampled from a single register. However, the consideration of this data did not enable a relevant set of categories to be identified through distributional learning. Though models which considered multidimensional acoustic data supported claim that additional acoustic dimensions can mitigate the ambiguity that is apparent in formant distributions (Swingley, 2009), such models still recovered too few categories. Comparisons of model performance across IDS and ADS did not indicate that the properties of IDS facilitated the use of distributional learning. These results reflected those of the acoustic analyses of IDS and ADS and provided further evidence against the hyperarticulation hypothesis. Models that were applied to registers with a lesser degree of overlap showed improved performance but this did not trivialise the infant learning task.

Chapter 5 operationalised the performance of these models by reporting the value of K, the number of categories that were identified, and measures of the pairwise F-score. The latter statistic is a compound measure of pairwise precision and recall and compares cases where the model judged a pair of tokens to be members of the same category with the status of that pair in the acoustic data. In the current set of models, the value of K was typically between six and eight for multidimensional acoustic data regardless of the speaker or register that the acoustic data was sampled from. As this value was less than the fifteen vowels of American English, these models predicted that infants would not be able to detect all of the relevant categories. The pairwise F-scores from across 100 runs of the clustering models for each speaker and register were typically less than .5. The current set of results are therefore comparable to values for K and the pairwise F-score that were observed in models that attempted to locate the number and identity of categories for a set of twelve American English vowels ($K = 8$, F-score = .52: Feldman et al., 2013; $K = 20$, F-score = .13: Antetomaso et al., 2016). Similar models have reported the V-measure which is akin to the F-score (V-measure = 53.9: Frank, Feldman, and Goldwater, 2014). Such statistics have also been reported for models which learnt five Japanese vowels ($K = 22$, F-score = .22: Antetomaso et al., 2016)) and nine Scottish English vowels ($K = 6$, F-score = .47: Mooney, 2015).

A further examination of these statistics indicated how the low value of K affected model performance. Pairwise F-scores were further broken down into pairwise measures of precision and recall. Across all of the models which were applied to acoustic data concerning different speakers, registers, and acoustic dimensions, pairwise precision was lower than pairwise recall. Low values for precision indicated that these models assigned many pairs of tokens to the same cluster which were distinct in the original data. This finding indicated that these models did not detect all of the relevant category boundaries. The observation that pairwise recall was high relative to pairwise precision should not be interpreted evidence of accurate performance. These values indicated that the models made a smaller proportion of false negatives relative to the number of false positives that they made. In other words, the models generally did not assign tokens of the same vowel category in the original data to separate clusters. Comparatively high values of pairwise recall followed from the low values of K as these models assigned a comparative small number of pairs of tokens to different clusters.

Models with simplifying assumptions Drawing contrasts between the performance of current set of clustering models and those where K is specified in advance indicates that this assumption greatly simplifies the learning task. The clustering models that were presented in this thesis and other comparable models viewed the estimation of the value of K as part of the learning task (Antetomaso et al., 2016; Feldman et al., 2013; Frank, Feldman, and Goldwater, 2014). Since these models were applied to American English data, they can be directly contrasted with a clustering model which attempted to identified exactly ten clusters from measures of F_1 and F_2 (Kornai, 1998). This model was applied to acoustic data which concerned a set of ten American English vowels (/i/, /ɪ/, /ɛ/, /æ/, /ɝ/, /ɑ/, /ʌ/, /ɔ/, /ʊ/, /u/: Peterson and Barney, 1952). This model was interpreted as providing support for the use of distributional learning as the mean formant values of the identified clusters closely resembled those which were reported in the original phonetic study. The observation that distributional learning could trivially identify the central tendencies of categories indicated that the task of locating an appropriate value for K represents a significant component of the infant learning task. Models which specify this value in advance do not resemble learning in infancy as learners do not have *a priori* knowledge of the number of distinctions in the inventory that they are acquiring. Because models where K is specified in advance have misrepresented the difficulty of the learning task, I argue that they do not provide relevant insights into the viability of distributional learning in infancy. The fact that model accuracy was solely assessed through comparisons of the central tendency also limited the relevance of these results. Measures of classification accuracy which indicate the proportion of vowel tokens which the model assigned to the correct category are required in order to understand how this model resolved cases of category overlap.

Models which have been applied to a subset of categories rather than an entire inventory have also overstated the viability of distributional learning, even when the value of K was not specified in advance. Comparisons between the current set of clustering models and those which have considered smaller sets of categories indicated that excluding categories enables successful categorisation. Moreover, models which have considered a subset of the relevant categories present the only cases where the value of K has been identified appropriately. Gaussian mixture models have successfully identified four categories from formant distributions which detailed comparable sets of front vowels from both Japanese IDS (/i, iː, ɛ, ɛː/) and American English IDS (/i, ɪ, eɪ, ɛ/: Vallabha et al., 2007). Models have also identified the value of K in formant data concerning two vowels sampled from Dutch IDS (/ɑ/, /aː/: Benders, 2013) and three vowels sampled from French IDS (/i/, /y/, /u/: Moeng, 2016). I propose two possible interpretations for the observation that clustering models are succeed when they are applied to subsets of categories rather than entire systems. On the one hand, these results may be interpreted as evidence that distributional learning can only be used to identify categories once learners have begun to process the vowel system. Such a claim would require infant learners to use other mechanisms in order to identify certain categories or to partition the system before they can draw inferences about category identity through distributional learning. This would negate one of the primary functions of this mechanism as the original proposal states that learners can identify the number of categories in a system by observing modes in the acoustic input. This approach would require a description of novel mechanisms which either enable the recognition of categories or that support and validate the generalisations from the distributional properties of the input. On the other hand, these results can also be interpreted as overstating the viability of distributional learning in infancy by excluding categories and therefore simplifying the infant learning task. Excluding categories in this way understates the degree of overlap between categories in acoustic space by eliminate confusable alternatives for each category. Under such an interpretation, models which learn four or fewer categories do not provide genuine insights into the viability of distributional learning in infancy.

Given the discussion above, it is unsurprising that models have successfully identified a subset of the categories in a system when the number of distinctions is specified in advance. Models of this type identified three clusters when they were applied to point vowel tokens that were sampled from American English IDS (/i/, /ɑ/ and /u/: de Boer and Kuhl, 2003; Kirchhoff and Schimmel, 2005). As in the model which considered ten vowels (Kornai, 1998), these results were interpreted as providing support for distributional learning as the mean formant values of the identified clusters closely corresponded to the genuine means of each point vowel. However, I instead view the accuracy of these models further highlighting the extent to which specifying the value of K and excluding vowels from the input can simplify the learning task. Point vowels are, by definition, acoustically dissimilar from one another and thus should present individual modes when they are considered in isolation. Two further studies have found that clustering models with a fixed number of categories were capable of identifying point vowels that were sampled from American English IDS with a high level of accuracy (Adriaans and Swingley, 2012, 2017). These models provided further insights into the viability of distributional learning as they showed poorer accuracy when they were applied to certain subsets of the system. The classification accuracy of models that were applied to tokens of /ɛ/, /æ/, and /ɑ/ was lower than that of the point vowel models. These models indicated that the distributional properties of the acoustic input did not enable accurate categorisation, even when a fixed number of isolated categories were learnt. The fact that cases of acoustic overlap hindered these models indicated that distributional learning may not be feasible mechanism even when it is supported with simplifying assumptions.

Hyperarticulation and multidimensional acoustic data In addition to highlighting cases where models have overstated the viability of distributional learning, I will now demonstrate that two limitations of previous learning models which learn an unknown number of categories have not led to an understatement of the viability of this mechanism (Antetomaso et al., 2016; Feldman et al., 2013; Frank, Feldman, and Goldwater, 2014). One limitation is that models of this type have only ever been applied to production data that was sampled from a single register. Register-specific differences in the discriminability of vowels could improve model performance regardless of whether they align with the predictions of the hyperarticulation hypothesis. The second limitation is that these models have exclusively identified categories on the basis of formant distributions. The use of multidimensional acoustic data could improve model performance as the consideration of additional acoustic dimensions may mitigate the ambiguity that is seen in formant data (Swingley, 2009). This claim is distinct from the previously refuted claim that the use of multidimensional acoustic data can reveal stronger effects of hyperarticulation in IDS (Eaves Jr. et al., 2016). The current set of models demonstrated that neither of these factors, even when combined, allowed for fifteen American English vowel categories to be identified from the statistical regularities of the acoustic input. Considering these factors did provide new insights into distributional learning in infancy as model performance across registers supported the results of the acoustic analyses and informed the selection of the measures of discriminability in this domain. The use of multidimensional acoustic data also facilitates the use of distributional learning as clustering models which were applied to high dimensional data had higher pairwise F-scores than formant models. However, these differences in model performance were numerically small and thus indicated that previous learning models have provided a genuine characterisation of the viability of distributional learning in infancy.

Given that the comparative acoustic analyses did not provide consistent evidence of contrast enhancement in IDS, it follows that clustering models for IDS data did not outperform the ADS clustering models. Register-specific differences in the pairwise F-score of these clustering models instead aligned with the measures of overlap that were reported in multidimensional acoustic space. For speakers ALI, CIN, and GAI who did not show a lesser degree of overlap in IDS relative to ADS, pairwise F-scores were greater in ADS than IDS. Conversely, pairwise F-scores were greater in IDS than ADS for speaker ANN who showed a lesser degree of overlap in this register. The magnitude of these effects was comparatively small: the largest of these effects was observed for speaker GAI who had a pairwise F-score of .608 in ADS and .476 in IDS. Though previous studies have not compared the performance of models across registers, this result is unsurprising since poor performance has been reported for models that were applied to Japanese and American English IDS data (Antetomaso et al., 2016). Each of these models estimated values of K that were much greater than the actual number of categories in each system. As the current set of models did not accurately categorise the system regardless of the register that they were applied to, the poor performance of previous models could not be attributed to the fact that these models have primarily considered ADS data. As discussed previously, the performance of clustering models provided support for the use of variance-sensitive measures as indicators of register-specific differences in discriminability. This evidence addressed the claim that measures in this domain have been overly dependent on researchers' intuitions about which properties facilitate learning (Eaves Jr. et al., 2016).

Chapter 4 described the realisation of the third formant, patterns of spectral change, and vowel duration in caregivers' speech and suggested that these dimensions supported the identification of category distinctions. The results of the clustering analyses indicated that these additional dimensions supported the use of distributional learning. In this way, they provided validation for the claim that multidimensional acoustic data can mitigate the ambiguity that is apparent in formant distributions (Swingley, 2009). Models that were applied to multidimensional data had higher F-scores than the formant distribution models across all combinations of registers and speakers. The direction of these results was consistent with the claim that previous studies which have solely considered measures of the first two formants may have understated the viability of distributional learning in infancy. However, the consideration of additional acoustic dimensions only had a small effect of the performance of the clustering models. The largest difference in model performance was observed for models that were applied to the ADS productions of speaker GAI. While the formant models had a pairwise F-score of .426, the multidimensional models had a pairwise F-score of 608. Though this represents a significant improvement in model performance, this comparison indicated that the previous approaches have not failed to identify a relevant set of categories because they have solely considered formant distributions.

In summary, the clustering models that were presented in chapter 5 indicated that distributional learning does not enable the identification of native language categories in infancy. I challenge the claim that previous models have providing support for the use of this mechanism in infancy because specifying the number of categories in advance or considering a subset of the inventory greatly simplifies this learning task. Successful categorisation in these cases should not be viewed as representative of the potential performance of an infant learner. Models which learn both the number and identity of the categories in the system have provided evidence that American English vowel categories cannot be identified by observing the statistical regularities of the acoustic input. The current analysis indicated that both register-specific differences in vowel quality and the use of multidimensional acoustic data improved the performance of clustering models. This indicates that previous models have understated the viability of this mechanism by solely considering formant data which was sampled from a single register. The size of these effects was numerically small, however, and cases where models which have not replicated learning in infancy cannot be attributed to either of these factors. The observation that distributional learning does not enable the identification of phonemic vowels of American English necessitates a reconsideration of the role of this mechanism in theories of perceptual development. The performance of clustering models which learn phonetic categories alongside lexical items (Feldman et al., 2013), semantic contexts (Frank, Feldman, and Goldwater, 2014), and phonological rules (Dillon, Dunbar, and Idsardi, 2013) highlight the possibility that distributional learning can function as a relevant mechanism when it is supported by infants' emergent knowledge of other linguistic units.

Logistic regressions A series of logistic regressions provided further insights into the distributional properties of the vowel production data. Rather than replicating learning in infancy, these models represented the extent to which ideal observer that was trained on the acoustic signal could predict the identity of the fifteen vowels. These models found that the acoustic data was an ambiguous predictor category identity, indicating that a learner with stronger inductive capabilities than an infant would still miscategorise vowel tokens. This analysis further supported the claim that multidimensional acoustic data facilitated distributional learning by indicating how each acoustic dimension contributed to the recognition of categories. The regression coefficients indicated that log vowel duration and patterns of spectral change tended to be stronger predictors of identity than measures of the third formant. The models also highlighted differences in the discriminability of individual categories and indicated that some low frequency categories were identified well. This result stands in contrast to the performance of the clustering models which generally did not recognise these categories. These models therefore indicate that the frequency of a category, in addition to its central tendency and variance, affected the viability of distributional learning in infancy.

Though logistic regressions illustrate the relationship between acoustic dimensions and category identity, they have not commonly been used as a method of assessing the distributional properties of the acoustic input. Regressions have demonstrated that formant measures strongly predicted the identity of a pair of American English distinctions (/i, ɪ/, /eɪ, ɛ/) while duration predicted the identity of two comparable Japanese distinctions (/i, iː/, /ɛ, ɛː/: Werker et al., 2007). The presence of these strong predictors was interpreted as evidence that distributional learning could successfully be applied to these samples of input. Multinomial logistic regressions have also been applied to nine American English vowels sampled from both IDS and ADS (McMurray et al., 2013). This analysis presented the observation that classification accuracy was greater in IDS than ADS as an indicator that a lack of global peripheralisation in IDS vowels should be interpreted as evidence that contrast enhancement did not occur in this register. Though this analysis did not make reference to the viability of distributional learning, the accuracy scores which were reported for individual categories indicated that peripheral vowels were more easily identified than internal vowels.

The classification accuracy of the logistic regressions that were applied to IDS data ranged from .639 to .708 while those of the regressions that were applied to ADS data ranged from .664 to .773. These accuracy measures are comparable to those that were reported for nine categories in American English (IDS: .578 − .716; ADS: .523 − .673: McMurray et al., 2013). No measures of classification accuracy were reported for the regressions that were applied to the American English or Japanese distinctions (Werker et al., 2007). These measures of performance indicated that an ideal observer is capable of accurately classifying the majority of tokens regardless of the speaker or register that they were sampled from. Since performance was not at ceiling, these regressions indicated that the distributional properties of the acoustic signal were ambiguous for all speakers and registers. I view logistic regressions as being similar to clustering models where K is specified in advance in terms of their relevance to the viability of distributional learning. Further to this, the relevance of these models is limited in cases where certain vowels have been excluded from the analysis. Regressions which have considered a single distinction are especially problematic in this regard (Werker et al., 2007). Nevertheless, these regressions have indicated that American English categories could be reliably discriminated once their identity is known. By pairing the results of clustering models and regressions, it is possible to determine the extent to which the presence of reliable predictors of category identity enables distributional learning. To my knowledge, the current analysis is the only case which has directly compared the performance of these two types of models. Indirect comparisons of this type can be drawn by comparing two separate analyses of phonemic vowel length in Japanese. While regressions indicated that vowel duration was a strong predictor for this distinction (Werker et al., 2007), an acoustic analysis of Japanese IDS found that vowel duration had a unimodal distribution which did not enable this distinction to be identified through distributional learning (Bion et al., 2013).

This analysis reported F-scores for each vowel category and thus supported for the claim that peripheral vowels are more easily discriminated than internal vowels (McMurray et al., 2013). The current set of regressions tended to have high F-scores for /i/, /u/ and /ɔɪ/ while low F-scores were observed for /ʌ/, /ɛ/, and /ɑ/. These results partially aligned with the study which originally observed this effect in a set of nine vowels of American English (McMurray et al., 2013). This regression analysis indicated that /i/, /æ/, and //oʊ/ were identified with a greater accuracy than /eɪ/ and /ɝ/. The observation of a high F-score for /ɔɪ/ in the current data set was consistent with the observation that acoustically extreme vowels are easily identified. Though /ɔɪ/ is not a point vowel, it had a low F_1 and F_2 at its onset and exhibited patterns of spectral change which were dissimilar from those of all other vowels in the system. The comparatively low F-scores for /oʊ/ and /æ/ meant that the current set of regressions did not entirely replicate the previous analysis. I propose that this difference in performance across studies again presents a case where the exclusion of categories affected model performance. The previous study reported greater accuracy for /oʊ/ as it was considered in the absence of three of its confusable alternatives, /ɔ/, /ʊ/ and /u/. The comparison of these two analyses again indicated that the exclusion of categories has lead to overstatements of the viability of statistical approaches in this domain.

Comparisons of the performance of the logistic regression and clustering models also highlighted cases where the frequency of specific categories in the input affected the distributional properties of the input. Though F-scores were comparatively high for the vowels /ɔɪ/ and /ʊ/, the clustering models did not identify individual clusters for either of these vowels because of their low frequency. Across the corpus as a whole, 0.6% of vowel tokens were tokens of /ɔɪ/ while 2.0% were tokens of /ʊ/. This effect was first identified in an analysis of phonemic vowel length in Japanese IDS (Bion et al., 2013). This study highlighted that low frequency categories may not present individual modes in the frequency distribution of the acoustic input even when their identity can be reliably predicted using regression analyses. Since only 6% of Japanese vowels are phonologically long, vowel duration has a unimodal distribution despite the existence of a reliable distinction. These two English vowels highlighted a case where the viability of distributional learning was affected by the frequency of a given category rather than its central tendency or variance. Despite this, assessments of the viability of distributional learning have typically not considered the frequency of individual categories to the same extent as measures of central tendency and variance. Clustering models which consider data sets where each category in data had the same frequency may therefore have overstated to discriminability of low frequency categories in the input (de Boer and Kuhl, 2003; Vallabha et al., 2007). Moreover, clustering models which have considered categories that vary in frequency have not explicitly considered the extent to which model performance is affected by this property of vowel categories in the input (Adriaans and Swingley, 2017; Benders, 2013; Feldman et al., 2013; Moeng, 2016; Mooney, 2015).

These logistic regressions confirmed the findings of the clustering models with regard to the effects that multidimensional acoustic data and register-specific differences in vowel quality had on model performance. The observation of significant regression coefficients for acoustic dimensions beyond the first two formants indicated that measures of the third formant, vowel duration, and patterns of spectral change were predictive of vowel identity across all speakers and registers. Comparisons of the number of categories which had a significant coefficient for each dimension indicated that log vowel duration and patterns of spectral change were more consistent predictors of identity than measures of the third formant. The current results therefore supported the claim that multidimensional acoustic data can mitigate the ambiguity of formant distributions (Swingley, 2009). Register-specific differences in the performance of the logistic regressions also aligned with the results of the comparative acoustic analyses of IDS and ADS. Numerical differences in the classification accuracy of the regressions aligned with the differences in multidimensional D(a) that were observed across registers. Classification accuracy was greater in ADS than IDS for speakers ALI, CIN, and GAI who exhibited a lack of enhancement in IDS. Conversely, classification accuracy was greater in IDS than ADS for speaker ANN which aligned with an effect of enhancement in IDS. Though these effects could not be verified through significance testing, the results indicated that register-specific differences in vowel quality affected the viability of distributional learning in infancy. Further to this, these results again supported the use of D(a) as a measure of discriminability.

Distributional learning in theories of perceptual attunement The observation that these models did not replicate learning in infancy necessitates a reconsideration of theories of perceptual attunement and the role that distributional learning has been attributed within them. These models demonstrated that the statistical regularities of caregivers' vowel production did not enable the identification of categories which correspond to the vowel phonemes of American English. Since these findings aligned with the observations of previous models which have attempted to learn an unknown number of categories (Antetomaso et al., 2016; Feldman et al., 2013; Frank, Feldman, and Goldwater, 2014), I have argued that such cases provide a genuine indicator of the viability of distributional learning and that models which have simplified the learning task have presented misleading views of this mechanism. The logistic regressions further indicated that the distributional properties of American English vowel categories were ambiguous even when models received explicit training. The clustering models and regression analyses indicated that model performance could be improved through the use of multidimensional acoustic data which was sampled from registers with a lesser degree of overlap. As these effects were comparatively small, the models predicted that the distributional properties of the input would lead learners to identify too few categories from the input even in optimal circumstances.

In order to reassess the viability of this mechanism, I will outline how distributional learning has been related to theories of perceptual development and evaluate the extent to which analyses of the statistical properties of IDS vowels have aligned with these theories. Developmental theories such as NLM (Kuhl, 1994; Kuhl et al., 1992, 2008) and PRIMIR (Werker and Curtin, 2005) state that exposure to the acoustic input is a primary factor that drives perceptual attunement in the first year of life. Current research has favoured distributional learning as an explicit account of how this experience with the acoustic signal modulates infants' perceptual sensitivities. Specifically, this mechanism enables infants to identify the number of categories in a system by observing the number of modes in the frequency distribution of the acoustic input. Learners can further fit Gaussian distributions to each mode in order to identify the limits of those categories. Distributional learning represents a common ground which has enabled further research into the use of statistical mechanisms in infants because it can be easily modelled and because it makes explicit predictions about the relevant properties of the input. Statistical learning paradigms, acoustic analyses of the input, and computational models of learning have all converged on the use of this mechanism as an explanatory account of the developmental patterns that have been observed through infant discrimination tasks. In order to align with the main set of distinctions which have been considered in infant discrimination tasks, studies which have evaluated the use of distributional learning have primarily considered phonemes to be the relevant set of phonetic categories in this learning task.

Although the establishment of this common ground has allowed for new insights into perceptual attunement, this position does not align well with developmental theories in this domain. These theories do not claim that phonemes are identified exclusively through the use of distributional learning. Because of this, it is not necessarily problematic that the current set of statistical models have failed to identify a set of clusters which correspond to the vowel phonemes of American English. Instead, I interpret these results as an indicator that researchers in this domain should reconsider both the mechanisms that infants use and the units that they identify. By doing so, current analyses of the acoustic input can contribute towards a better characterisation of the role that distributional learning plays in perceptual attunement. In terms of mechanisms, current developmental theories state that the identification of phonetic categories stands in a mutually beneficial relationship with infants' emergent knowledge of lexical items, phonotactics, and prosodic structures. In terms of the units to be learnt, a recent critique of perceptual attunement has argued that phonemes should not be viewed as the goal of distributional learning (Dillon, Dunbar, and Idsardi, 2013). This critique has claimed that current approaches have implicitly described perceptual attunement as a two-stage approach. In the first stage, learners use statistical mechanisms to identify individual phones which are then grouped into phonologically relevant units in the second stage. Identifying phonemes as perceptually invariant units in the first stage is undesirable as it would render the learners' sensitivity to allophonic variation. The authors have therefore proposed a single-stage approach in which the learner simultaneously identifies phones and the rules which relate between them. In order to address issues regarding the mechanisms and goals of this task, I will now evaluate a series of modelling studies which have demonstrated how a reformed approach to distributional learning may function.

Clustering models have provided evidence that phonetic categories can be identified successfully when distributional learning is combined with the identification of other linguistic units. Models which simultaneously learnt lexical items and vowel categories outperformed those that solely learnt vowel categories (Feldman et al., 2013; Frank, Feldman, and Goldwater, 2014). These lexical-distributional models identified phonetic categories and lexical items on the basis of vowel tokens which consisted of a pair of formant measures and a word frame. As an example, this model should label a token with low values for F_1 and F_2 in the /b_k/ frame as being an instance of the vowel /ʊ/ in the word ⟨book⟩. Models of this type successfully identified twelve clusters with a pairwise F-score of .92 from American English ADS data which detailed a set of twelve monophthongs. By contrast, models which solely learnt phonetic categories identified eight clusters with a pairwise F-score of .52. The improved performance of these models provided evidence in principle learning category- and word-level inferences can bootstrap each other as highlighted in theories of perceptual attunement. Prior beliefs about category identity enabled the identification of vowel minimal pairs while beliefs about lexical identity resolved cases where vowel categories overlapped. Models which can additionally consider semantic contexts have extended this effect of bootstrapping (Frank, Feldman, and Goldwater, 2014). Models which learnt vowel categories, lexical items, and their semantic contexts further outperformed those which learnt vowels categories and lexical items (Frank, Feldman, and Goldwater, 2014). These models received the context or activity that a word occurred in in addition to formant measures and word frames. Lexical ambiguities could be resolved on this basis as well as through differences in vowel quality. For example, the words ⟨book⟩ and ⟨bike⟩ could be distinguished as the former is associated with reading while the latter is associated with outdoors.

Though these models have provided evidence in principle of a positive interaction between phonetic categories and other linguistic units, these results cannot be extended to learning in infancy as these models had abilities which exceeded those of infant learners between the age of 0;6 and 0;8. Each model learnt a potentially unlimited number of lexical items from input which fully specified the identity of consonants. At this point of development, infants have a limited receptive vocabulary and have a poorer discriminatory capabilities for consonants than vowels. American English infants aged 0;6 have been shown to recognise a small number of common nouns but did not detect segment-level mispronunciations in these forms until the age of 0;11 (Bergelson and Swingley, 2012, 2017). Experimental tasks have also indicated that infants also form a protolexicon that consists of both words and non-words around this age. French infants aged 0;11 distinguished high-frequency disyllables from those with a low frequency regardless of whether they corresponded words in their native language (Ngon et al., 2013). In order to determine the validity of pairing distributional learning with word learning mechanisms, it is necessary to implement models with capabilities that more closely resemble infants' capabilities in word learning at this age. The idea that consonantal identity can be used to bootstrap the recognition of vowel categories is also contrary to the results of discrimination tasks. Such tasks have indicated that infants show language-specific perceptual behaviours for vowel distinctions before they do so with consonantal distinctions (Tsuji and Cristia, 2013). The validity of this assumption has been tested by applying learning models to data where the identity of the consonants in word frames is limited Frank, Feldman, and Goldwater (2014). Specifically, models were presented with word frames where consonants were only specified for manner rather than voicing, place, and manner. Even with this limited information, models which learnt phonetic categories and lexical items outperformed those that learnt phonetic categories in isolation. These results have therefore indicated that even a partial knowledge of the adult lexicon could have a positive effect on distributional learning in infancy.

Models have also addressed the observation that this mechanism does not identify clusters which correspond to American English vowel phonemes by reconsidering the goals of perceptual attunement. Previous modelling work has illustrated this type of approach using acoustic data which was sampled from Inuktitut IDS (Dillon, Dunbar, and Idsardi, 2013). This language features six phones that correspond to three phonemes, /i/, /a/ and /u/, and their lowered allophones, [e], [ɑ], and [o] which occur before uvular consonants. The authors have proposed that distributional learning can either serve as the first stage of a two-stage approach or that it can be combined with other mechanisms in a single-stage approach. In the former case, the goal of this mechanism would be to identify the individual six phones of Inuktitut so that they can be grouped together in order to establish phonemic categories. In the latter case, the distributional properties of the input can be considered alongside phonological variables in order to enable the identification of six phones and a rule of lowering that relates the phonemes to their allophones.

These Inuktitut clustering models provided evidence against the two-stage approach as they tended to recover three clusters rather than six allophones when they were presented with formant distributions from Inuktitut IDS (Dillon, Dunbar, and Idsardi, 2013). In cases where the model identified six clusters, these solutions did not necessarily capture the allophonic distinctions in height. For example, these models identified a spurious distinction in advancement within tokens of /u/ and [o]. The authors described such results as problematic as the two-stage approach requires phones in the first stage to be identified accurately so that learners make generalisations about the environments that condition them at the next stage. The results of previous clustering models can be considered as aligning with this observation as they have generally identified a number of clusters which was smaller than the number of phonemes in the system (Feldman et al., 2013; Frank, Feldman, and Goldwater, 2014; Mooney, 2015). The outputs of these models cannot be further grouped together in the second stage in order to identify phonemic categories. Clustering models which were applied to Japanese and American English IDS data have identified a number of clusters which was greater than the number of phonemes in the system (Antetomaso et al., 2016). As this study did not discuss the identity of these clusters, it was not possible to determine the extent to which these clusters corresponded to relevant phones of Japanese or American English.

The authors further presented a clustering model which learnt both phones and allophonic rules as support for the single-stage approach (Dillon, Dunbar, and Idsardi, 2013). This computational model learnt an unknown number of pairs of categories and a set of rules which linked them from Inuktitut vowel tokens which consisted of formant measures and an indicator of whether the token occurred in the conditioning environment. This model aimed to identify the central tendency and variance of each of the three vowels, /i/, /a/ and /u/, as well as a rule which indicated the difference between the central tendencies of each phoneme and their lowered allophones. This allophonic model identified an appropriate rule of lowering and outperformed the model which was not sensitive to the conditioning environments. This difference in performance was interpreted as evidence that a single-stage approach is a better characterisation of perceptual attunement and therefore justifies a reconsideration of the goals and mechanisms of this learning task. However, the implications that this single-stage approach has for the use of distributional learning in infancy must be interpreted by considering the extent to which these models resemble the learning task that infants face and the cognitive skills that are available to them at this point in development.

The vowel system of Inuktitut and the allophonic patterns within it present an idealised case which illustrates how an emergent knowledge of phonotactics could bootstrap an infant's perception of phonetic categories. However, I will argue that the simplicity of this system and the model's assumptions mean that the current set of results should not be interpreted as evidence that the limitations of distributional learning can be addressed by pairing it with phonotactic information. Three vowel systems are one of the simplest typologically common configurations and clustering models have demonstrated that distributional learning can identify these three point vowels (Adriaans and Swingley, 2012, 2017; de Boer and Kuhl, 2003; Kirchhoff and Schimmel, 2005). The rule of lowering in this system is also comparatively simple as it occurs in a single environment, namely before uvulars. The presence of such consonants also has a consistent phonetic consequence on each vowel in the system. This process of tracking conditioning environments and identifying their phonetic consequences is more complex for /t/ in American English, for example. American English learners would have to identify lexical stress, syllable position, and prosodic structure as factors that condition the occurrence of a set of phonetically dissimilar allophones including [tʰ], [ɾ] and [t̚]. In essence, the single-stage approach predicts that learners must identify multiple prosodic, coarticulatory or assimilatory processes in order to map many phones to the phonetically invariant realisation of each native language category. The authors also acknowledge that their models do not incorporate this potentially complex process of identifying conditioning environments into the learning task. The tokens in the input to these model were marked with a binary coding which indicated whether vowel token occurred in that environment and had only to establish the phonetic consequence that is associated with it.

In order to more fully assess the viability of distributional learning in infancy, further modelling work must closely consider both the mechanisms that are associated with this task and the categories that those mechanisms identify. Though previous modelling work has indicated that distributional learning can be used to identify phonetic categories in this domain, they have often made simplifying assumptions which limit their relevance to the infant learning task. As such, these models cannot be interpreted as providing support for the use of distributional learning outside of laboratory contexts. The current set of learning models indicate that distributional learning does not permit the identification of categories which correspond to the phonemic vowel categories of American English. This was the case even when multidimensional data which was sampled from a register with a lesser degree of overlap was considered. The poor performance of these models provides a genuine indicator of the viability of this mechanism and thus necessitates a reconsideration of theories of perceptual attunement in this domain. Though a series of innovative models have provided evidence that the identification of phonetic categories can be bootstrapped using infants' emergent knowledge of lexical items, semantic contexts, and phonotactics, these approaches have still made simplifying assumptions regarding infants' abilities to identify and process these additional units of linguistic structure. Because of this, current theories of perceptual attunement and the use of distributional learning within these approaches must be validated through a series of models which more greatly resemble the cognitive abilities of infants within the first year of life. Future studies which consider the viability of this mechanism must provide evidence a noisy and partial knowledge of other levels of linguistic structure can be used to establish and validate a set of phonetic categories which are relevant to later stages of phonological development.

6.4 Conclusions and future directions

This thesis was motivated by two observations which concern the emergence of language-specific behaviours in the perception of American English vowel categories. The first observation is that caregivers alter their vowel production when they address infant learners in a similar way across the world's languages. Functionalist explanations of the features of this register state that these adaptations promote language learning in infancy. The second observation is the shift towards language-specific perceptual behaviour which infants exhibit within the first year of life. Theories of perceptual attunement state that exposure to the statistical regularities of the input drives this learning process. The acoustic analyses from chapters 3 and 4 challenged functionalist approaches to IDS vowel production as the properties of this register did not facilitate the recognition and processing of vowel distinctions. The learning models from chapter 5 challenged current assumptions about the mechanisms behind this learning task as they indicated that native language distinctions cannot be identified on the basis of statistical regularities of the acoustic input. The current section will present the conclusions of this thesis in a broader theoretical context and outline the further issues in this domain which are prompted by these conclusions.

Hyperarticulation in IDS The hyperarticulation hypothesis states that the properties of IDS vowel production indicate an intention to facilitate the identification of native language vowel distinctions in infancy (Bernstein Ratner, 1984; Kuhl et al., 1997). This claim is part of a larger functionalist approach which claims that the features of this register are shaped by caregivers' intention to promote language learning. Such an account would support the observation that IDS bears the same features across the majority of well-reported languages (Cristia, 2013; Saint-Georges et al., 2013; Soderstrom, 2007). The fact that comparative studies have not provided unanimous evidence of enhancement in IDS is a central issue which was addressed by this thesis. Conflicting results in this domain have primarily consisted of comparisons of the area of the vowel space. Though this is the most commonly reported measures of discriminability in this domain, it is a poor indicator of the distributional properties of the input. As well as highlighting that area of the vowel space only describes the formant values of a subset of categories, I have argued that this measure's inability to capture register-specific differences in variance is its main flaw. The acoustic analyses revealed a lack of enhancement in IDS by applying variance-sensitive measures of discriminability to multidimensional acoustic data for a full set of American English vowel categories. This analysis provided evidence of a larger vowel space and greater dispersion in IDS relative to ADS and thus indicated the central tendencies of all categories within the system were consistent with enhancement. Despite this, the greater within-category variance that was observed in IDS meant that categories did not show a lesser degree of overlap in IDS relative to ADS. Greater within-category variance in IDS relative to ADS has commonly been reported in this domain (American English, Russian and Swedish: Kuhl et al., 1997; American English: Cristia and Seidl, 2014, Kirchhoff and Schimmel, 2005, McMurray et al., 2013; Dutch: Benders, 2013; Japanese: Miyazawa et al., 2017) and variance-sensitive measures have only indicated a lack of enhancement in IDS (Cristia and Seidl, 2014; Miyazawa et al., 2017). In addition to advocating for the use of variance-sensitive measures in this domain, the current set of analyses indicated that multidimensional acoustic data did not consistently provide stronger evidence of enhancement in IDS than formant distributions did. I therefore refuted the claim that additional acoustic dimensions must be considered to assess the discriminability of IDS and ADS (Eaves Jr. et al., 2016). Previous formant analyses which have not provide evidence of enhancement in IDS should therefore be viewed as relevant to the hyperarticulation hypothesis.

By assessing multiple measures of discriminability, the current analyses clarified the mixed evidence that acoustic studies have presented with regard to the hyperarticulation hypothesis. As vowel space expansion has been demonstrated to be a poor indicator of discriminability, I have argued that conflicting observation for this statistic should not be interpreted as having strong implications for the hyperarticulation hypothesis. Existing samples of acoustic data from IDS and ADS must be reanalysed using further measures of discrimination in order to better describe how these measures inform the hyperarticulation hypothesis. Given that the current study has indicated that more nuanced measures of discriminability do not align with effects of expansion, this type of renewed analysis may provide a more coherent set of results in this domain. Since the current analysis reported greater dispersion for majority of IDS distinctions relative to ADS, these results need to be reconciled with previous cases where measures of dispersion have indicated a lack of enhancement in IDS (Bohn, 2013; Cristia and Seidl, 2014). As previous studies have only considered a limited number of distinctions, they have not provided conclusive evidence against the hyperarticulation hypothesis. Without assessing a broader range of distinctions, it cannot be determined whether whether this lack of enhancement was representative of the discriminability of all IDS distinctions or whether these effects were false negatives in a system which broadly exhibited an effect of enhancement. The observation of greater dispersion in IDS also influenced the interpretation of the measures of overlap which indicated a lack of enhancement in IDS. This observation, paired with the fact that a greater degree of overlap was observed in IDS relative to ADS in some cases, indicated that it would not be sufficient to merely refuted the hypothesis. Refuting this claim would predict that the distributional properties of each register should be comparable. Because this was not the claim, I elected to interpret the effects of dispersion and variance as having distinct causes. I proposed that the effect of dispersion indicated that caregivers attempted to enhance IDS distinctions and that this intention did not succeed because of external factors which resulted in greater within-category variance in IDS. Although the heightened variance of IDS vowels is a well reported effect in this domain, current studies have yet to establish and quantify the casual factors behind the greater variance of acoustic dimensions in this register. Further analyses must therefore closely consider aspects of the input such as pitch and sentence length which have previously been described as more variable in IDS than ADS (Cristia and Seidl, 2014; Soderstrom, 2007) For example, such investigations would have to confirm that the variance of F_0 was greater in IDS than ADS before identifying the extent to which register-specific differences in the variance of acoustic dimensions can be attributed to the variance of pitch. Without such evidence, the negative effects of heightened variance in IDS which have been presented in this thesis would remain unmotivated.

The current set of comparative acoustic analyses focussed on the relative discriminability of vowels in each register. Because of this, these analyses did not fully address two alternative proposals regarding the properties of IDS vowel production. One proposal claims that the acoustic properties of IDS enable the communication of positive affect rather than serving to enhance distinctions (Benders, 2013). The current analyses did not support this account of IDS vowel production as it did not observe a consistent effect of formant raising in IDS. This approach has, however, received support from studies which have adopted novel methodologies in this domain. Specifically, articulatory measures (Kalashnikova, Carignan, and Burnham, 2017), listeners' ratings of maternal affect (Benders, 2016) and measures of voice quality (Miyazawa et al., 2017) have all indicated that IDS has greater positive affect than ADS. Future analyses must therefore account for differences in maternal affect when assessing the properties of IDS vowel production. The second proposal states that the characteristics of IDS vowel production are better explained as a result of the differences in the prosodic properties of each register (McMurray et al., 2013). The corpus which I analysed consisted of caregivers' spontaneous speech and thus vowel tokens occurred across a variety of prosodic positions and word types in each register. Because of this, it was not possible to establish how the acoustic properties of vowel categories differed across registers independently of these effects of prosody. Given that studies have demonstrated that controlling for these prosodic factors can reduce the magnitude of register-specific differences in vowel production (McMurray et al., 2013; Wang, Seidl, and Cristia, 2015), the current analysis may have overstated the extent to which the central tendency and variance of vowels differed as a result of the addressee that caregivers interacted with.

Further considerations of spontaneous IDS and ADS are required in order to gain insights into the differences in the prosodic structure of each of these registers. Such a description of each register must then be supplemented with controlled phonetic studies that indicate how the acoustic properties of IDS and ADS vowels differ within a specific context. Such studies should ideally apply measures of discriminability to acoustic data that has been sampled from these contexts as current studies have solely considered differences in the identity of individual categories.

Though the acoustic properties of IDS vowel categories were not consistent with the hyperarticulation hypothesis, this lack of enhancement does not negate the claim that other properties of this register are shaped by an intention to promote language learning in infancy. Though the properties of the acoustic signal did not enable the use of distributional learning (Maye, Werker, and Gerken, 2002), theories of perceptual development maintain that the identification of phonetic categories stands in a mutually beneficial relationship with the identification of lexical items and phonotactic rules. Because of this, features of IDS which enable the acquisition of other levels of linguistic structure can be viewed as indirectly enabling perceptual attunement in infancy. The observation that infants preferentially attend to this register as opposed to ADS, for example, can therefore be said to facilitate the identification and processing of phonetic categories (Cooper and Aslin, 1990; Fernald and Kuhl, 1987). The observation that IDS captures and holds infant attention more effectively than ADS has been linked with the slower speech rate, higher pitch and larger pitch excursions that can be observed in caregivers' speech. These alterations to the acoustic properties of IDS are therefore functionally motivated. Cases where infants who were presented with IDS input outperform those presented with ADS in tasks involving word segmentation (Thiessen, Hill, and Saffran, 2005), word recognition (Singh et al., 2009; Song, Demuth, and Morgan, 2010) and the identification of phrasal boundaries (Kemler Nelson et al., 1989) have highlighted positive effects associated with infant attention. Further to this, I argue that the general observation that IDS is highly variable can be associated with the observation that high variability input is required to ensure the emergence of robust vowel category perception in infancy (Cristia and Seidl, 2014; Miyazawa et al., 2017). It is important to note that this effect of variance can only be relevant once learners have established the set of the vowel distinctions in their native language: such effects must further be distinguished from cases where variance in acoustic dimensions negatively affects the use of distributional learning.

Distributional learning in infancy Distributional learning is an mechanism which indicates how exposure to the statistical properties of acoustic signal alters infants' perception of speech sounds within the first year of life (Maye, Werker, and Gerken, 2002). Specifically, exposure to bimodal input results in a continued sensitivity to a native distinction while exposure to unimodal input attenuates that sensitivity. As such, this mechanism describes the maintenance of native distinctions and the loss of all others which infants exhibit within their first year of life (Tsuji and Cristia, 2013). Though it aligns closely with theories which state that experience with the acoustic input drives perceptual attunement (Kuhl, 1994; Kuhl et al., 1992, 2008; Werker and Curtin, 2005), the viability of distributional learning crucially depends on an assumption the frequency distribution of the acoustic input presents learners with a mode for each category in their native language. I have argued previous learning models have overstated the viability of this mechanism in infancy. Models which have replicated infant learning have only succeeded because they simplified this task by specifying the number of categories to be learnt, considering a subset of the relevant system, or making both of these assumptions. By contrast, models which have attempted to learn both the number and identity of categories from an entire inventory have shown much poorer performance (Antetomaso et al., 2016; Feldman et al., 2013; Frank, Feldman, and Goldwater, 2014; Mooney, 2015). By applying clustering models and logistic regressions to multidimensional data that was sampled from both IDS and ADS, chapter 5 indicated that distributional learning was not sufficient to identify categories which corresponded to the phonemic vowels of American English. In doing so, it indicated that the poor performance of models with fewer assumptions have presented a more valid characterisation of the viability of distributional learning. As performance remained poor when these models were applied to multidimensional acoustic data from both IDS and ADS, I have demonstrated that models which have solely considered formant distributions from a single register have not understated the viability of statistical approaches. These results necessitate a reform of both the mechanisms and the goals of perceptual attunement. Renewed approaches to this task must align more closely with theories of perceptual development which propose that distributional learning works alongside other learning mechanisms in order to establish categories which are relevant to later phonological development.

By applying learning models to acoustic data from a naturalistic corpus of IDS and ADS, I have demonstrated that learners who solely consider the statistical properties of the acoustic input would identify between six and eight clusters rather than the fifteen phonemic vowels of American English. These results contrasted with the accurate performance of models which made simplifying assumptions about the learning task. Given that models where number of categories was known in advance have trivially identified the mean formant values for ten categories, the current set of models demonstrated that identifying the number of categories in a system is a complex stage of this learning task (Kornai, 1998). Similarly, the accurate performance of models which have considered a limited number of categories have reduced the complexity of the task by eliminating cases of overlap which infant learners would have to resolve (Benders, 2013; Moeng, 2016; Vallabha et al., 2007). In order to provide an accurate characterisation of the learning task, future modelling should attempt to learn both the number and identity of phonetic categories in an entire inventory. Additionally, these models should consider multidimensional acoustic data that is sampled from IDS. The current analysis indicate that multidimensional acoustic data mitigates the case of overlap that have been observed in formant distributions (Swingley, 2009). Though these additional acoustic dimensions resulted in a significant improvement in model performance, this effect was numerically small and did not enable the identification of all fifteen categories. Similarly, the use of IDS data did not enable successful categorisation as register-specific differences in the performance of models aligned with the lack of enhancement that was observed in this register. Chapter 5 also presented a series of logistic regressions which indicated that multidimensional acoustic data from both IDS and ADS provided learners with ambiguous predictors of category identity in American English. Such approaches have not commonly been adopted in this domain, despite the fact that they indicate how individual acoustic dimensions enable the identification of each category in a system. Given that these regressions additionally highlighted that acoustically extreme vowels are more easily discriminated and that the frequency of vowels affects their discriminability, I have advocated for further uses of this statistical technique in this domain. Differences in the frequency of categories were first described in (Bion et al., 2013) and have received minimal attention in considerations of distributional learning. This aspect of the input should be considered closely when assessing the results of clustering models as these models may not detect low-frequency categories even when their acoustic properties are strongly predictive of a distinction.

As stated previously, these clustering models necessitate a reform of both the mechanisms and goals that have been associated with perceptual attunement. By doing so, computational modelling approaches have described implementations of distributional learning which are closer to current theories of perceptual development. With regard to learning mechanisms, clustering models have provided evidence in principle that there phonetic categories can be successfully identified alongside lexical items and semantic contexts (Feldman et al., 2013; Frank, Feldman, and Goldwater, 2014). The assumptions of these models prompts further studies in this domain: although these models made reasonable assumptions regarding the use of distributional learning in infancy, their abilities to identify lexical information exceeded those of infant learners. Further models which make realistic assumptions about infants' vocabulary are required in order to fully evaluate the benefits of bootstrapping phonetic categories in this way. Similarly, further modelling work must provide a better characterisation of the goals of this learning task. A model that was applied to Inuktitut data has provided insights into how the identification of phonotactic rules can ensure the establishment of a phonological relevant set of categories (Dillon, Dunbar, and Idsardi, 2013). However, I view the vowel system of Inuktitut as a simplistic case which provides a simplified illustration of this type of approach. Further work is needed to establish the extent to which a statistical learner can identify the environments which condition these allophonic rules. Models must also indicate the extent to which a sensitivity to phonotactics enables accurate categorisation in systems with a larger number of phonemes or with more complex cases of allophony.

Reconsidering the goals of perceptual attunement also prompts an reconsideration of the current experimental evidence regarding perceptual attunement and the use of distributional learning in infancy. The results of discrimination tasks have generally indicated that broad-based discrimination at birth gives way to language-specific behaviour through the loss of non-native distinctions. Distributional learning is apt to describe these patterns of development as infants' behaviour in unimodal conditions have been consistent with the loss of non-native distinctions. However, this mechanism has shown a poorer alignment with cases where certain distinctions are poorly discriminated at birth or where learners' sensitivity to both native and non-native distinctions increases throughout development (Mazuka, Hasegawa, and Tsuji, 2014). Further discrimination tasks are required to more fully establish the sets of distinctions which infants easily discriminate at birth, those which they maintain, and those for which sensitivity increases throughout development. By establishing the timeline of perceptual development more closely, it is possible to gain further insights into the mechanisms that learners can exploit in this learning task. Distributional learning has been favoured as an early explanatory mechanism as it solely requires infants to observe the statistical regularities of the acoustic input. By contrast, cases of perceptual attunement in later development can be explained by incorporating infants' emergent knowledge of lexical items and phonotactic patterns. Further experimental tasks are also required to further establish the availability and relevance of distributional learning with regard to vowel perception in infancy. While this mechanism has replicated across a range of consonantal contrasts in both infant and adult learners, only a single study has exhibited this effect for vowel distinctions. Exposure to distributional information affected how Dutch infants between the ages of 0;2 and 0;3 discriminated the English vowels /ɛ/ and /æ/ (Wanrooij, Boersma, and Benders, 2015). By contrast, exposure to distributional information did not affect how Canadian English infants aged 0;8 perceived vowel distinctions (/ɛ/, /ɛː/: Pons et al., 2006a; /ɪ/, /e/: Pons et al., 2006b). Establishing that distributional learning is a robust mechanism would justify the central role that it has been attributed in theories of perceptual attunement. Further to this, exposure to a bimodal distributions has only ever been demonstrated that increase perceptual sensitivity in adult learners (Escudero, Benders, and Wanrooij, 2011; Gulian, Escudero, and Boersma, 2007; Wanrooij, 2015). Demonstrating such an effect with infants would extend the potential uses for this mechanism in a developmental context.

Closing remarks This thesis challenged current views regarding the hyperarticulation of vowels in IDS, the viability of distributional learning in infancy, and the interaction of these two aspects of the emergence of language-specific perceptual behaviours in infancy. This thesis presented a detailed comparative analysis of the acoustic properties of vowel production in IDS and ADS. New empirical data did not support the claim that the properties of IDS are shaped by an intention to promote the identification and processing of vowel categories in infancy. The use of variance-sensitive measures revealed that previous assessments of the hyperarticulation hypothesis have failed to observe that the greater within-category variance of vowels in IDS has a negative effect of the discriminability of distinctions in this register. IDS categories therefore exhibited a comparable or greater degree of overlap than their ADS counterparts, hindering the use of distributional learning in infancy. Further studies in this domain must therefore adopt variance-sensitive measures and provide explanations for this well-reported property of IDS vowel production. This thesis further applied a series of clustering models and logistic regressions to the acoustic data which I sampled from each register. Measures of model performance indicated that distributional learning did not enable learners to identify the phonetic categories of American English. Instead, clustering models indicated that exposure to the acoustic properties of the input would reduce learners' sensitivity to native distinctions while logistic regressions indicated that acoustic variables were ambiguous predictors of category identity. These models revealed that previous computational methods have overstated the viability of this mechanism by making simplifying assumptions about the learning task that infants face. These results necessitate a reconsideration of the role that this mechanism plays in perceptual attunement in infancy. Specifically, it prompts a holistic approach where infants' emergent knowledge of lexical items and phonotactic processes can bootstrap the identification of phonetic categories in infancy. Further modelling work must closely replicate learning in infancy and consider the extent to which these factors can be combined with distributional learning in order to ensure the identification of a phonologically relevant set of perceptual categories.

References

Adriaans, F. and D. Swingley (2012). "Distributional learning of vowel categories is supported by prosody in infant-directed speech". In: *Proceedings of the 34th Annual Conference of the Cognitive Science Society*. Ed. by N. Miyake, D. Peebles, and R. P. Cooper, pp. 72–77.

— (2017). "Prosodic exaggeration within infant-directed speech: Consequences for vowel learnability". In: *Journal of the Acoustical Society of America* 141.5, pp. 3070–3078.

Antetomaso, S., N. H. Feldman, K. Miyazawa, M. Elsner, K. Hitczenko, and R. Mazuka (2016). "Modeling phonetic category learning from natural acoustic data". Presentation given at the Boston University Conference on Language Development.

Benders, T. (2013). "Nature's distributional-learning experiment: Infants' input, infants' perception, and computational modeling". PhD thesis. University of Amsterdam.

— (2016). "Emotion-related explanations of the vowel variability in infant-directed speech". In: *Proceedings of the Sixteenth Australasian International Conference on Speech Science and Technology*. Ed. by C. Carignan and M. D. Tyler, pp. 233–236.

Bergelson, E. and D. Swingley (2012). "At 6–9 months, human infants know the meanings of many common nouns". In: *Proceedings of the National Academy of Sciences* 109, pp. 3253–3258.

— (2017). "Young Infants' Word Comprehension Given An Unfamiliar Talker or Altered Pronunciations". In: *Child Development:* Advance online publication.

Bernstein Ratner, N. (1984). "Patterns of vowel modification in mother–child speech". In: *Journal of Child Language* 11.3, pp. 557–578.

Best, C. T. (1994). "The development of speech perception: the transition from speech sounds to spoken words". In: ed. by J. C. Goodman and H. C. Nusbaum. Cambridge, MA: MIT Press. Chap. The emergence of native-language phonological influences in infants: a Perceptual Assimilation Model, pp. 167–224.

Best, C. T., G. W. Roberts, and N. M. Sithole (1988). "Examination of perception reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants". In: *Journal of Experimental Psychology: Human Perception and Performance* 14, pp. 345–360.

Bion, R., K. Miyazawa, H. Kikuchi, and R. Mazuka (2013). "Learning Phonemic Vowel Length from Naturalistic Recordings of Japanese Infant-Directed Speech". In: *PLOS One* 8.2, pp. 1–5.

Boersma, P. and D. Weenink (2018). *Praat: doing phonetics by computer*. Version 6.0.40, accessed at `http://www.praat.org/`.

Bohn, O.-S. (2013). "Acoustic characteristics of Danish infant-directed speech". In: *Proceedings of the Meetings on Acoustics* 19, pp. 1–7.

Burnham, D., C. Kitamura, and U. Vollmer-Conna (2002). "What's new, pussycat? On talking to babies and animals." In: *Science* 296.5572, p. 1435.

Burnham, E. B., E. A. Wieland, M. V. Kondaurova, J. D. McAuley, T. R. Bergeson, and L. C. Dilley (2015). "Phonetic modification of vowel space in storybook speech to infants up to two years of age". In: *Journal of Speech, Language, and Hearing Research* 58, pp. 241–253.

Capel, D., J. H. De Bree, M. A. De Klerk, A. O. Kerkhoff, and F. N. K. Wijnen (2011). "Distributional cues affect phonetic discrimination in Dutch infants". In: *Sound and Sounds: Studies Presented to M. E. H. Scouten on the Occasion of His 65th Birthday*. Ed. by W. Zonneveld, H. Quené, and W. Heeren, pp. 33–43.

Cardillo, G. C. (2010). “Predicting the predictors: Individual differences in longitudinal relationships between infant phonetic perception, toddler vocabulary, and preschooler language and phonological awareness”. PhD thesis. University of Washington.

Caselli, M. C., E. Bates, P. Casadio, J. Fenson, L. Fenson, L. Sanderl, and J. Weir (1995). “A cross-linguistic study of early lexical development”. In: *Cognitive Development* 10, pp. 159–199.

Chambers, K., K. Onishi, and C. Fisher (2003). “Infants learning phonotactic regularities from brief auditory experience”. In: *Cognition* 87, B69–B77.

Cheour, M., K. Alho, R. Ceponiené, K. Reinikainen, K. Sainio, M. Pohjavuori, O. Aaltonen, and R. Näätanen (1998). “Maturation of mismatch negativity in infants”. In: *International Journal of Psychophysiology* 29.2, pp. 217–226.

Clark, M. J. and J. M. Hillenbrand (2007). “Quality of back vowels before /r/ in the American Inland North”. In: *Journal of the International Phonetic Association* 37.3, pp. 275–291.

Cooper, R. P. and R. N. Aslin (1990). “Preferences for infant-directed speech in the first month after birth”. In: *Child Development* 61, pp. 1584–1595.

Cristia, A. (2013). “Input to language: The phonetics and perception of infant-directed speech”. In: *Language and Linguistics Compass* 7, pp. 157–170.

Cristia, A. and A. Seidl (2014). “The hyperarticulation hypothesis of infant-directed speech”. In: *Journal of Child Language* 41.4, pp. 913–934.

Cristia, A., G. L. McGuire, A. Seidl, and A. L. Francis (2011). “Effects of the distribution of acoustic cues on infants’ perception of speech sounds”. In: *Journal of Phonetics* 39.3, pp. 388–402.

de Boer, B. and P. K. Kuhl (2003). “Investigating the role of infant-directed speech with a computer model”. In: *Acoustics Research Letters Online* 4, pp. 2238–2248.

Dillon, B., E. Dunbar, and W. Idsardi (2013). “A single-stage approach to learning phonological categories: Insights from Inuktitut”. In: *Cognitive Science* 37.2, pp. 344–377.

Dodane, C. and J. Al-Tamimi (2007). “An acoustic comparison of vowel systems in adult-directed speech and child-directed speech: Evidence from French, English & Japanese”. In: *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbruecken: University of Saarbruecken, pp. 1573–1576.

Eaves Jr., B. S., N. H. Feldman, T. L. Griffiths, and P. Shafto (2016). “Infant-directed speech is consistent with teaching”. In: *Psychological Review* 123.7, pp. 758–771.

Eckert, P. and W. Labov (2017). “Phonetics, phonology and social meaning”. In: *Journal of Sociolinguistics* 21.4, pp. 467–496.

Eimas, P. D., E. R. Siqueland, P. W. Jusczyk, and J. Vigorito (1971). “Speech perception in infants”. In: *Science* 171.3968, pp. 303–306.

Elsner, M. and K. Ito (2017). “An Automatically Aligner Corpus of Child-directed Speech”. In: *Proceedings of Interspeech*. Stockholm, Sweden.

Englund, K. T. (2005). “Infant Directed Speech in Natural Interaction – Norwegian Vowel Quantity and Quality”. In: *Journal of Psycholinguistic Research* 34.3, pp. 259–280.

Englund, K. T. and D. Behne (2006). “Changes in Infant Directed Speech in the First Six Months”. In: *Infant and Child Development* 15, pp. 139–160.

Escudero, P., T. Benders, and K. Wanrooij (2011). “Enhanced bimodal distributions facilitate the learning of second language vowels”. In: *Journal of the Acoustical Society of America* 130.1, pp. 416–428.

Feldman, N. H., T. L. Griffiths, S. Goldwater, and J. L. Morgan (2013). "A Role for the Developing Lexicon in Phonetic Category Acquisition". In: *Psychological Review* 120.4, pp. 751–778.

Fernald, A. and P. K. Kuhl (1987). "Acoustic determinants of infant preference for motherese speech". In: *Infant Behavior and Development* 10, pp. 279–293.

Fernald, A., T. Taeschner, J. Dunn, M. Paposek, B. deBoysson Bardies, and I. Fukui (1989). "A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants". In: *Journal of Child Language* 16.3, pp. 477–501.

Fraley, C. and A. E. Raftery (2006). *MCLUST Version 3 for R: Normal Mixture Modeling and Model-Based Clustering*.

Frank, S., N. H. Feldman, and S. Goldwater (2014). "Weak semantic context helps phonetic learning in a model of infant language acquisition". In: *Proceedings of the 52nd Annual Meeting of the Association of Computational Linguistics*, pp. 1073–1083.

Fruehwald, J. (2013). "The Phonological Influence on Phonetic Change". PhD thesis. University of Pennsylvania.

Garcia, C. (2012). "A simple procedure for the comparison of covariance matrices". In: *BMC Evolutionary Biology* 12, pp. 1–17.

Gomez, R. L. and L. A. Gerken (1999). "Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge". In: *Cognition* 70, pp. 109–135.

Grieser, D. and P. K. Kuhl (1989). "Categorization of speech by infants: support for speech-sound prototypes". In: *Developmental Psychology* 25, pp. 577–588.

Gulian, M., P. Escudero, and P. Boersma (2007). "Supervision hampers distributional learning of vowel contrasts". In: *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbrücken: University of Saarbrücken, pp. 1893–1896.

Hall-Lew, L., M. Eiswirth, M.-C. Valentinsson, and W. Cotter (2017). "Duke University Press". In: *Speech in the Western States, Volume 2: The Mountain West*. Ed. by V. Fridland, A. Wassink, T. Kendall, and B. E. Evans, pp. 59–82.

Hartman, K. M., N. Bernstein Ratner, and R. S. Newman (2016). "Infant-directed speech (IDS) vowel clarity and language outcomes". In: *Journal of Child Language* 44.5, pp. 1140–1162.

Hillenbrand, J., L. A. Getty, M. J. Clark, and K. Wheeler (1995). "Acoustic characteristics of American English vowels". In: *Journal of the Acoustical Society of America* 97, pp. 3099–3111.

Hillenbrand, J. M. (2013). "Static and dynamic approaches to vowel perception". In: *Vowel Inherent Spectral Change*. Ed. by G. S. Morrison and P. F. Assmann. Vol. Springer-Verlag. Heidelberg, pp. 9–30.

Jansson-Verkasalo, E., T. Ruusuvirta, M. Huotilainen, P. Alku, E. Kushnerenko, K. Suominen, S. Rytky, M. Luotonen, T. Kaukola, U. Tolonen, and M. Hallman (2010). "Atypical perceptual narrowing in prematurely born infants is associated with compromised language acquisition at 2 years of age". In: *BMC Neuroscience* 11.1, pp. 1–7.

Kalashnikova, M., C. Carignan, and D. Burnham (2017). "The origins of babytalk: smiling, teaching or social convergence?" In: *Royal Society Open Science* 4.170306, pp. 1–11.

Kemler Nelson, D., K. Hirsh-Pasek, P. W. Jusczyk, and K. W. Cassidy (1989). "How prosodic cues in motherese might assist language learning". In: *Journal of Child Language* 16, pp. 55–68.

Kirby, J. (2014). "Acquisition of covert contrast: an unsupervised learning approach". In: *Proceedings of the 46th Annual Meeting of the Chicago Linguistic Society*. Ed. by R. Baglini, T. Grinsell, J. Keane, A. R. Singerman, and J. Thomas. Vol. 46. 2, pp. 111–125.

Kirchhoff, K. and S. Schimmel (2005). "Statistical properties of infant-directed versus adult-directed speech: Insights from speech recognition". In: *Journal of the Acoustical Society of America* 117.4, pp. 2238–2246.

Ko, E.-S. and M. Soderstrom (2013). "Additive Effects of Lengthening on the Utterance-Final Word in Child-Directed Speech". In: *Journal of Speech, Language, and Hearing Research* 56, pp. 364–371.

Kornai, A. (1998). "Analytic models in phonology". In: *The organization of phonology: Constraints, levels and representations*. Ed. by J. Durand and B. Laks. Oxford, UK: Oxford University Press, pp. 395–418.

Kuhl, P. K. (1979). "Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories". In: *Journal of the Acoustical Society of America* 66, pp. 1668–1679.

— (1983). "Perception of auditory equivalence classes for speech in early infancy". In: *Infant Behavior and Development* 6, pp. 263–285.

— (1991). "Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not". In: *Perception & Psychophysics* 50, pp. 93–107.

— (1994). "Learning and representation in speech and language". In: *Current Opinion in Neurobiology* 4.6, pp. 812–822.

Kuhl, P. K., K. A. Williams, F. Lacerda, K. N. Stevens, and B. Lindblom (1992). "Linguistic experience alters phonetic perception in infants by 6 months of age". In: *Science* 31.255, pp. 606–608.

Kuhl, P. K., J. E. Andruski, I. A. Chistovich, E. V. Kozhevenikova, V. L. Ryskina, E. I. Stolyarova, U. Sundberg, and F. Lacerda (1997). "Cross-language analysis of phonetic units in language addressed to infants". In: *Science* 277, pp. 684–686.

Kuhl, P. K., E. Stevens, A. Hayashi, T. Deguchi, S. Kiritani, and P. Iverson (2006). "Infants show a facilitation effect for native language phonetic perception between 6 and 12 months". In: *Developmental Science* 9.2, F13–21.

Kuhl, P. K., B. T. Conboy, S. Coffey-Corina, D. Padden, M. Rivera-Gaxiola, and T. Nelson (2008). "Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e)". In: *Philosophical Transactions of the Royal Society B* 363, pp. 979–1000.

Labov, W., S. Ash, and C. Boberg (2005). *Atlas of North American English: Phonetics, phonology and sound change*. Berlin; New York: Mouton de Gruyter.

Labov, W., I. Rosenfelder, and J. Fruehwald (2013). "One Hundred Years of Sound Change in Philadelphia: Linear Incrementation, Reversal, and Reanalysis". In: *Language* 89.1, pp. 30–65.

Ladefoged, P. and I. Maddieson (1996). *The Sounds of the World's Languages*. Blackwell.

Lam, C. and C. Kitamura (2010). "Maternal Interactions with a Hearing and Hearing-Impaired Twin: Similarities and Differences in Speech Input, Interaction Quality and Word Production". In: *Journal of Speech, Language, and Hearing Research* 53, pp. 543–555.

— (2012). "Mommy, speak clearly: induced hearing loss shapes vowel hyperarticulation". In: *Developmental Science* 15.2, pp. 212–221.

Li, P. and Y. Shirai (2000). *The acquisition of lexical and grammatical aspect.* New York, NY: Mouton de Gruyter.

Liljencrants, J. and B. Lindblom (1972). "Numerical simulation of vowel quality systems: The role of perceptual contrast". In: *Language* 48, pp. 839–862.

Lindblom, B. (1986). "On the origin and purpose of discreteness and invariance in sound patterns". In: *Invariance and Variability in Speech Processes.* Ed. by J. S. Perkell and D. H. Klatt. Lawrence Erlbaum, pp. 493–510.

Lisker, L. and A. S. Abramson (1964). "Cross-language study of voicing in initial stops: acoustical measurements". In: *Word* 20, pp. 384–422.

Liu, H.-M., P. K. Kuhl, and F.-M. Tsao (2003). "An association between mother's speech clarity and infants' speech discrimination skills". In: *Developmental Science* 6.3, pp. 1–10.

Liu, H.-M., F.-M. Tsao, and P. K. Kuhl (2009). "Age-related Changes in Acoustic Modifications of Mandarin Maternal Speech to Preverbal Infants and Five-Year-Old Children: A Longitudinal Study". In: *Journal of Child Language* 36.4, pp. 909–922.

MacWhinney, B. (2000). *The CHILDES Project: tools for analyzing talk.* Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Mani, N. and K. Plunkett (2010). "Twelve-month-olds know their cups from their keps and tups". In: *Psychological Science* 15.5, pp. 445–470.

Marean, G., L. Werner, and P. K. Kuhl (1992). "Vowel categorisation by very young infants". In: *Developmental Psychology* 28.3, pp. 396–405.

Martin, A., T. Schatz, M. Versteegh, K. Miyazawa, R. Mazuka, E. Dupoux, and A. Cristia (2015). "Mothers Speak Less Clearly to Infants Than to Adults: A Comprehensive Test of the Hyperarticulation Hypothesis". In: *Psychological Science* 26.3, pp. 341–347.

Maye, J., D. J. Weiss, and R. N. Aslin (2008). "Statistical phonetic learning in infants: facilitation and feature generalization". In: *Developmental Science* 11.1, pp. 122–134.

Maye, J., J. F. Werker, and L. A. Gerken (2002). "Infant sensitivity to distributional information can affect phonetic discrimination". In: *Cognition* 82, pp. 101–111.

Mazuka, R., M. Hasegawa, and S. Tsuji (2014). "Development of Non-Native Vowel Discrimination: Improvement Without Exposure". In: *Developmental Psychobiology* 56.2, pp. 192–209.

Mazuka, R., Y. Igarashi, and K. Nishikawa (2006). "Input for learning Japanese: RIKEN Japanese mother-infant conversation corpus". In: *IEICE Technical Report* 16, pp. 11–15.

McMurray, B., R. N. Aslin, and J. C. Toscano (2009). "Statistical learning of phonetic categories: Insights from a computational approach". In: *Developmental Science* 12.3, pp. 369–378.

McMurray, B., K. A. Kovack-Lesh, D. Goodwin, and W. McEchron (2013). "Infant directed speech and the development of speech perception: Enhancing development or an unintended consequence?" In: *Cognition* 129.2, pp. 362–378.

Miyawaki, K., J. J. Jenkins, W. Strange, A. M. Liberman, R. Verbrugge, and O. Fujimura (1975). "An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and Engli". In: *Perception & Psychophysics* 18.5, pp. 331–340.

Miyazawa, K., T. Shinya, A. Martin, H. Kikuchi, and R. Mazuka (2017). "Vowels in infant-directed speech: More breathy and more variable, but not clearer". In: *Cognition* 166, pp. 84–93.

Moeng, E. (2016). "Comparing the distributional learnability of vowels, glides, and stops". In: *Proceedings of the 40th Annual Boston University Conference on Language Development.* Ed. by J. Scott and D. Waughtal.

Mooney, S. (2015). "Extending a North American English category learner to a non-standard variety: Categorizing vowels across speech styles in Glaswegian English". In: *Proceedings of the 18th International Congress of Phonetic Sciences.* Ed. by The Scottish Consortium for ICPhS 2015.

Narayan, C. R., J. F. Werker, and P. S. Beddor (2010). "The interaction between acoustic salience and language experience in developmental speech perception: evidence from nasal place discrimination". In: *Developmental Science* 13.3, pp. 407–420.

Nearey, T. M. and P. F. Assmann (1986). "Modeling the role of vowel inherent spectral change in vowel identification". In: *Journal of the Acoustical Society of America* 80, pp. 1297–1308.

Neel, A. T. (2008). "Vowel space characteristics and vowel identification accuracy". In: *Journal of Speech, Language, and Hearing Research* 51, pp. 574–585.

Newman, R. S., S. A. Clouse, and J. L. Burnham (2001). "The perceptual consequences of within-talker variability in fricative production". In: *Journal of the Acoustical Society of America* 109.3, pp. 1181–1196.

Ngon, C., A. Martin, E. Dupoux, D. Cabrol, M. Dutat, and S. Peperkamp (2013). "(Non)words, (non)words, (non)words: evidence for a protolexicon during the first year of life". In: *Developmental Science* 16.1, pp. 24–34.

Ohala, J. (1980). "The acoustic origin of the smile". In: *Journal of the Acoustical Society of America* 68, S33.

— (1984). "An ethological perspective on common cross-language utilisation of F0 of voice". In: *Phonetica* 41, pp. 1–16.

Pegg, J. E. and J. F. Werker (1997). "Adult and infant perception of two English phones". In: *Journal of the Acoustical Society of America* 102, pp. 3742–3753.

Peterson, G. E. and H. L. Barney (1952). "Control methods used in a study of the vowels". In: *Journal of the Acoustical Society of America* 24, pp. 175–184.

Polka, L. and O.-S. Bohn (1996). "A cross-language comparison of vowel perception in English-learning and German-learning infants". In: *Journal of the Acoustical Society of America* 100.1, pp. 577–592.

— (2011). "Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development". In: *Journal of Phonetics* 39, pp. 467–478.

Polka, L., C. Colantonio, and M. Sundara (2001). "A cross-language comparison of /d/–/ð/ perception: Evidence for a new developmental pattern". In: *Journal of the Acoustical Society of America* 109.5, pp. 2190–2201.

Polka, L. and J. F. Werker (1994). "Developmental change in perception of nonnative vowel contrasts". In: *Journal of Experimental Psychology: Human Perception and Performance* 20, pp. 421–435.

Pons, F., R. Mugitani, S. Amano, and J. F. Werker (2006a). "Distributional learning in vowel length distinctions by 6-month-old English infants". Poster presented at the annual meeting of the XVth Biennial International Conference on Infant Studies.

Pons, F., L. Sabourin, J. C. Cady, and J. F. Werker (2006b). "Distributional learning of vowel distinctions by 8-month-old English infants". Poster presented at the 28th Annual Conference of the Cognitive Science Society.

R Core Team (2013). *R: A Language and Environment for Statistical Computing.* Accessed at http://www.R-project.org/.

Rosenfelder, I., J. Fruehwald, K. Evanini, and J. Yuan (2011). *FAVE (Forced Alignment and Vowel Extraction) Program Suite*. Accessed at http://fave.ling.upenn.edu.

Rost, G. C. and B. McMurray (2009). "Speaker variability augments phonological processing in early word learning". In: *Developmental Science* 12.2, pp. 339–349.

Saffran, J. R., R. N. Aslin, and E. L. Newport (1996). "Statistical learning by 8-month-old infants". In: *Science* 20, pp. 421–435.

Saint-Georges, C., M. Chetouani, R. Cassel, F. Apicella, A. Mahdhaoui, F. Muratori, M-C. Laznik, and D. Cohen (2013). "Motherese in interaction: at the cross-road of emotion and cognition? (A systematic review)". In: *PLOS One* 8.10, pp. 1–17.

Sebastián-Gallés, N. and L. Bosch (2009). "Developmental shift in the discrimination of vowel contrasts in bilingual infants: is the distributional account all there is to it?" In: *Developmental Science* 12.6, pp. 874–887.

Singh, L., S. Nestor, C. Parikh, and A. Yull (2009). "Influences of infant-directed speech on early world recognition". In: *Infancy* 14, pp. 654–666.

Soderstrom, M. (2007). "Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants". In: *Developmental Review* 27, pp. 501–532.

Sonderegger, M., M. Bane, and P. Graff (2017). "The medium-term dynamics of accents on reality television". In: *Language* 93.3, pp. 598–640.

Song, J. Y., K. Demuth, and J. L. Morgan (2010). "Effects of the acoustic properties of infant-directed speech on infant word recognition". In: *Journal of the Acoustical Society of America* 128, pp. 389–400.

Stager, C. L. and J. F. Werker (1997). "Infants listen for more phonetic detail in speech perception than in word-learning tasks". In: *Nature* 388, pp. 381–382.

Sundara, M. and A. Scutellaro (2011). "Rhythmic distance between languages affects the development of speech perception in bilingual infants". In: *Journal of Phonetics* 39.4, pp. 505–513.

Sundberg, U. and F. Lacerda (1999). "Voice Onset Time in Speech to Infants and Adults". In: *Phonetica* 56, pp. 186–199.

Swingley, D. (2009). "Contributions of infant word learning to language development". In: *Philosophical Transactions of the Royal Society B: Biological Sciences*. Vol. 364, pp. 3617–3632.

Swoboda, P. J., J. Kass, P. A. Morse, and L. A. Leavitt (1978). "Memory factors in vowel discrimination of normal and at-risk infants". In: *Child Development* 49.2, pp. 332–339.

Thiessen, E. D. (2011). "When Variability Matters More Than Meaning: The Effect of Lexical Forms on Use of Phonemic Contrasts". In: *Developmental Psychology* 47.5, pp. 1448–1458.

Thiessen, E. D., E. A. Hill, and J. R. Saffran (2005). "Infant-directed speech facilitates word segmentation". In: *Infancy* 7.1, pp. 53–71.

Trainor, L. J. and R. N. Desjardins (2002). "Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels". In: *Psychonomic Bulletin & Review* 9, pp. 335–340.

Trehub, S. E. (1973). "Infants' senstivity to vowel and tonal contrasts". In: *Developmental Psychology* 9.1, pp. 91–96.

Tsuji, S. and A. Cristia (2013). "Fifty years of infant vowel discrimination research: what have we learned?" In: *Journal of the Phonetic Society of Japan* 17.3, pp. 1–11.

Uther, M., M. A. Knoll, and D. Burnham (2007). “Do you speak E-NG-L-I-SH? A comparison of foreign- and infant-directed speech”. In: *Speech Communcation* 49, pp. 2–7.

Vallabha, G. K., J. L. McClelland, F. Pons, J. F. Werker, and S. Amano (2007). “Unsupervised learning of vowel categories from infant-directed speech”. In: *Proceedings of the National Academy of Sciences* 104, pp. 13273–13278.

Wang, Y., A. Seidl, and A. Cristia (2015). “Acoustic-phonetic differences between infant- and adult-directed speech: the role of stress and utterance position”. In: *Journal of Child Language* 42.4, pp. 821–842.

Wanrooij, K. (2015). “Distributional learning of vowel categories in infants and adults”. PhD thesis. University of Amsterdam.

Wanrooij, K., P. Boersma, and T. Benders (2015). “Observed effects of “distributional learning” may not relation to the number of peaks. A test of dispersion as a confounding factor”. In: *Frontiers in Psychology* 6.1341, pp. 1–13.

Werker, J. F. and S. Curtin (2005). “PRIMIR: a developmental framework of infant speech processing”. In: *Language Learning and Development* 1.2, pp. 197–234.

Werker, J. F. and R. C. Tees (1984). “Cross-language speech perception: Evidence for perceptual reorganization during the first year of life”. In: *Infant Behavior and Development* 7, pp. 49–63.

Werker, J. F., F. Pons, C. Dietrich, C. Kajikawa, L. Fais, and S. Amano (2007). “Infant-directed speech supports phonetic category learning in English and Japanese”. In: *Cognition* 103, pp. 147–162.

Wieland, E. A., E. B. Burnham, M. Kondaurova, T. R. Bergeson, and L. C. Dilley (2015). “Vowel Space Characteristics of Speech Directed to Children With and Without Hearing Loss”. In: *Journal of Speech, Language, and Hearing Research* 58, pp. 254–267.

Xu, N., D. Burnham, C. Kitamura, and U. Vollmer-Conna (2013). “Vowel hyperarticulation in parrot-, dog-and infant-directed speech”. In: *Anthrozoos* 26.3, pp. 373–380.

Xu Rattanasone, N., D. Burnham, and R. G. Reilly (2013). “Tone and vowel enhancement in Cantonese infant-directed speech at 3, 6, 9, and 12 months of age”. In: *Journal of Phonetics* 41.5, pp. 332–343.

Yoshida, K., F. Pons, J. Maye, and J. F. Werker (2010). “Distributional phonetic learning at 10 months of age”. In: *Infancy* 15.4, pp. 420–433.

Yuan, J. and M. Liberman (2008). “Speaker identification on the SCOTUS corpus”. In: *Proceedings of Acoustics 2008*, pp. 5687–5690.

— (2011). “Automatic detection of ‘g-dropping’ in American English using forced alignment”. In: *Proceedings of 2011 IEEE Automatic Speech Recognition and Understanding Workshop*, pp. 490–493.

A Confusion matrices

The following tables provide further detail concerning the results of the multinomial logistic regressions that were presented in 5.4.1. This set of tables provides exact counts for the confusion matrices that are displayed in figures 44 and 45. Tables 40, 41, 42, and 43 provide the exact counts for the models that were applied to the IDS productions of ALI, ANN, CIN, and GAI respectively. Tables 44, 45, 46, and 47 do so for the models that were applied to their ADS productions. Exact counts for these correspondences were used to calculate the classification accuracy for the regressions as well as measures of the F-score, precision, and recall for individual categories.

		predicted category														
		ɑ	æ	ʌ	ɔ	aʊ	ɑɪ	ɛ	ɝ	eɪ	ɪ	i	oʊ	ɔɪ	ʊ	u
actual category	ɑ	90	–	13	49	1	12	–	–	–	–	–	28	–	6	1
	æ	1	78	9	1	49	1	31	3	20	6	–	1	–	–	–
	ʌ	10	2	113	2	8	25	2	8	1	–	–	11	–	18	–
	ɔ	45	–	4	114	–	15	–	–	–	–	–	21	–	1	–
	aʊ	4	38	6	–	121	1	20	–	4	3	–	3	–	–	–
	ɑɪ	11	–	14	6	2	164	–	–	–	–	–	3	–	–	–
	ɛ	1	37	11	–	16	–	58	13	25	27	6	2	–	3	1
	ɝ	–	–	3	–	–	–	18	139	1	28	–	–	–	4	7
	eɪ	–	12	5	–	–	1	16	–	146	11	9	–	–	–	–
	ɪ	–	1	–	–	1	–	22	31	5	105	24	–	–	–	11
	i	–	–	–	–	–	–	2	1	6	18	173	–	–	–	–
	oʊ	19	–	10	15	7	2	–	–	–	–	–	128	2	16	1
	ɔɪ	2	–	–	–	–	–	–	–	–	–	–	1	184	4	9
	ʊ	1	–	14	–	–	–	1	6	–	–	–	8	7	152	11
	u	–	–	–	–	1	–	2	9	3	11	2	–	11	10	151

Table 40: A confusion matrix indicating how the predictions of the multinomial logistic regression applied to ALI's IDS data aligned with the actual identity of vowels in the resampled data.

		predicted category														
		ɑ	æ	ʌ	ɔ	aʊ	ɑɪ	ɛ	ɝ	eɪ	ɪ	i	oʊ	ɔɪ	ʊ	u
actual category	ɑ	81	–	50	36	6	14	–	–	–	–	–	7	–	6	–
	æ	–	114	3	–	19	4	35	6	15	4	–	–	–	–	–
	ʌ	45	4	79	2	15	29	2	–	–	–	–	16	–	8	–
	ɔ	22	–	9	133	4	1	–	–	–	–	–	9	8	9	5
	aʊ	14	28	21	5	128	–	3	–	–	–	–	1	–	–	–
	ɑɪ	8	2	21	–	5	164	–	–	–	–	–	–	–	–	–
	ɛ	1	30	10	–	1	2	96	8	7	41	2	2	–	–	–
	ɝ	1	1	–	–	–	1	3	151	2	23	–	5	–	6	7
	eɪ	–	5	–	–	–	4	4	1	181	4	1	–	–	–	–
	ɪ	–	1	–	–	–	–	31	19	4	108	26	1	–	–	10
	i	–	–	–	–	–	–	–	–	6	16	178	–	–	–	–
	oʊ	5	2	10	14	11	–	1	2	–	–	–	144	–	7	4
	ɔɪ	–	–	–	–	–	–	–	–	–	–	–	–	196	4	–
	ʊ	5	–	–	12	–	–	–	8	–	–	–	7	3	160	5
	u	–	1	–	2	–	–	1	6	–	5	1	5	–	11	168

Table 41: A confusion matrix indicating how the predictions of the multinomial logistic regression applied to ANN's IDS data aligned with the actual identity of vowels in the resampled data.

		predicted category														
		ɑ	æ	ʌ	ɔ	aʊ	ɑɪ	ɛ	ɝ	eɪ	ɪ	i	oʊ	ɔɪ	ʊ	u
actual category	ɑ	112	1	32	22	12	12	–	–	–	–	–	8	–	1	–
	æ	–	135	8	–	26	1	24	1	5	–	–	–	–	–	–
	ʌ	26	5	96	1	19	16	–	9	–	–	–	19	–	8	1
	ɔ	28	–	–	135	–	–	–	–	–	–	–	3	–	30	4
	aʊ	12	23	10	–	132	6	7	1	–	–	–	9	–	–	–
	ɑɪ	15	1	11	–	3	169	–	1	–	–	–	–	–	–	–
	ɛ	–	17	4	–	10	–	123	8	19	19	–	–	–	–	–
	ɝ	–	3	8	–	–	–	11	137	5	27	–	6	–	–	3
	eɪ	–	6	1	–	–	1	14	2	144	16	16	–	–	–	–
	ɪ	–	–	–	–	–	–	16	35	10	120	19	–	–	–	–
	i	–	–	–	–	–	–	–	–	10	13	177	–	–	–	–
	oʊ	9	–	12	7	13	–	1	4	1	–	–	133	–	17	3
	ɔɪ	–	–	–	1	–	–	–	–	–	–	–	–	195	1	3
	ʊ	3	–	7	31	–	–	–	–	–	–	–	18	5	132	4
	u	–	–	–	1	–	–	–	5	–	–	–	–	5	4	185

Table 42: A confusion matrix indicating how the predictions of the multinomial logistic regression applied to CIN's IDS data aligned with the actual identity of vowels in the resampled data.

		predicted category														
		ɑ	æ	ʌ	ɔ	aʊ	ɑɪ	ɛ	ɝ	eɪ	ɪ	i	oʊ	ɔɪ	ʊ	u
actual category	ɑ	125	1	25	5	12	9	–	–	–	–	–	13	–	10	–
	æ	–	153	2	–	13	6	16	2	8	–	–	–	–	–	–
	ʌ	31	4	98	1	10	21	2	10	–	–	–	14	–	9	–
	ɔ	27	–	3	86	2	–	–	–	–	–	–	18	10	28	26
	aʊ	8	13	11	–	145	6	7	1	–	–	–	8	–	1	–
	ɑɪ	9	7	27	–	1	156	–	–	–	–	–	–	–	–	–
	ɛ	1	7	9	–	7	2	123	18	18	13	–	1	–	1	–
	ɝ	–	–	6	1	1	–	16	148	2	16	–	–	–	7	3
	eɪ	–	7	–	–	–	–	16	2	165	10	–	–	–	–	–
	ɪ	–	–	–	–	–	–	22	16	6	137	12	–	–	–	7
	i	–	–	–	–	–	–	–	–	2	10	188	–	–	–	–
	oʊ	7	–	7	18	11	–	1	4	–	–	–	132	6	11	3
	ɔɪ	–	–	–	19	–	–	–	–	–	–	–	1	179	1	–
	ʊ	8	–	12	49	–	1	–	1	–	–	–	11	2	116	–
	u	–	–	–	5	–	–	–	5	–	5	5	–	7	3	170

Table 43: A confusion matrix indicating how the predictions of the multinomial logistic regression applied to GAI's IDS data aligned with the actual identity of vowels in the resampled data.

		predicted category														
		ɑ	æ	ʌ	ɔ	aʊ	ɑɪ	ɛ	ɝ	eɪ	ɪ	i	oʊ	ɔɪ	ʊ	u
actual category	ɑ	129	–	4	50	–	12	–	–	–	–	–	5	–	–	–
	æ	–	121	8	–	38	3	18	–	11	–	–	1	–	–	–
	ʌ	5	1	124	–	4	17	8	13	2	1	–	23	–	2	–
	ɔ	38	–	–	145	1	5	–	–	–	–	–	2	1	8	–
	aʊ	–	38	2	–	148	4	3	–	–	–	–	5	–	–	–
	ɑɪ	11	1	15	2	1	163	–	–	2	–	–	5	–	–	–
	ɛ	–	26	9	–	1	3	114	8	6	22	6	1	–	–	4
	ɝ	–	–	5	1	–	–	4	136	1	5	–	16	–	15	17
	eɪ	–	9	3	–	4	1	7	4	170	2	–	–	–	–	–
	ɪ	–	–	–	–	–	–	16	10	11	116	30	–	–	–	17
	i	–	–	–	–	–	–	3	–	1	29	160	–	–	–	7
	oʊ	3	4	25	6	7	3	3	4	–	–	–	145	–	–	–
	ɔɪ	–	–	–	–	–	–	–	–	–	–	–	–	200	–	–
	ʊ	–	–	1	4	–	–	–	8	–	–	–	–	1	180	6
	u	–	–	–	4	–	–	–	–	–	12	23	–	–	15	146

Table 44: A confusion matrix indicating how the predictions of the multinomial logistic regression applied to ALI's ADS data aligned with the actual identity of vowels in the resampled data.

		predicted category														
		ɑ	æ	ʌ	ɔ	aʊ	ɑɪ	ɛ	ɝ	eɪ	ɪ	i	oʊ	ɔɪ	ʊ	u
actual category	ɑ	104	–	19	38	16	12	–	2	–	–	–	6	–	3	–
	æ	1	124	3	–	14	7	41	2	3	2	–	3	–	–	–
	ʌ	27	7	51	13	37	13	7	5	1	–	–	33	2	4	–
	ɔ	35	–	3	124	5	7	–	–	–	–	–	10	–	16	–
	aʊ	22	9	50	2	89	11	2	–	–	–	–	14	–	1	–
	ɑɪ	5	5	10	–	11	168	–	–	–	–	–	1	–	–	–
	ɛ	–	44	4	1	3	1	110	5	14	4	–	14	–	–	–
	ɝ	–	1	1	1	–	–	16	133	14	8	–	12	3	2	9
	eɪ	–	1	1	–	–	–	14	10	153	18	2	1	–	–	–
	ɪ	–	–	–	–	–	–	5	8	22	136	19	–	–	–	10
	i	–	–	–	–	–	–	–	–	3	21	167	–	–	–	9
	oʊ	5	9	15	9	11	–	11	14	–	–	–	101	–	24	1
	ɔɪ	–	–	–	–	–	–	–	2	1	–	–	–	194	3	–
	ʊ	4	–	–	10	–	–	–	–	–	–	–	–	5	175	6
	u	–	–	–	2	–	–	–	4	–	6	17	1	4	4	162

Table 45: A confusion matrix indicating how the predictions of the multinomial logistic regression applied to ANN's ADS data aligned with the actual identity of vowels in the resampled data.

		predicted category														
		ɑ	æ	ʌ	ɔ	aʊ	ɑɪ	ɛ	ɝ	eɪ	ɪ	i	oʊ	ɔɪ	ʊ	u
actual category	ɑ	106	6	16	40	4	21	–	–	–	–	–	6	–	1	–
	æ	5	128	1	–	36	11	16	–	3	–	–	–	–	–	0
	ʌ	28	7	100	2	8	12	11	3	1	–	–	25	–	2	1
	ɔ	39	–	9	121	1	–	–	–	–	–	–	12	–	18	–
	aʊ	8	30	13	–	138	4	5	–	–	–	–	2	–	–	–
	ɑɪ	12	8	11	–	4	165	–	–	–	–	–	–	–	–	–
	ɛ	1	7	13	–	5	1	130	1	16	25	–	1	–	–	–
	ɝ	–	–	4	–	–	–	12	148	1	14	–	5	–	–	16
	eɪ	–	4	–	–	–	6	11	–	155	23	1	–	–	–	–
	ɪ	–	–	–	–	–	–	18	20	23	115	13	–	–	–	11
	i	–	–	–	–	–	–	–	–	2	17	179	–	–	–	2
	oʊ	2	–	18	15	7	–	1	10	–	–	–	123	7	16	1
	ɔɪ	–	–	1	–	–	–	–	–	–	–	–	7	189	3	–
	ʊ	–	–	–	9	–	–	–	3	–	–	–	4	10	174	–
	u	–	–	–	–	–	–	–	9	–	10	5	1	–	4	171

Table 46: A confusion matrix indicating how the predictions of the multinomial logistic regression applied to CIN's ADS data aligned with the actual identity of vowels in the resampled data.

		predicted category														
		ɑ	æ	ʌ	ɔ	aʊ	ɑɪ	ɛ	ɝ	eɪ	ɪ	i	oʊ	ɔɪ	ʊ	u
actual category	ɑ	142	3	21	19	–	14	–	–	–	–	–	1	–	–	–
	æ	1	143	2	–	22	8	23	–	1	–	–	–	–	–	–
	ʌ	25	9	97	14	8	14	11	5	2	–	–	14	–	1	–
	ɔ	29	–	7	143	–	–	–	–	–	–	–	5	9	7	–
	aʊ	–	19	13	–	141	1	11	–	–	–	–	15	–	–	–
	ɑɪ	13	10	14	–	–	160	1	–	2	–	–	–	–	–	–
	ɛ	–	15	8	–	4	1	140	8	13	4	–	7	–	–	–
	ɝ	–	–	9	–	–	–	8	155	–	6	–	9	–	4	9
	eɪ	–	–	3	–	–	1	13	1	173	9	–	–	–	–	–
	ɪ	–	–	–	–	–	–	2	7	9	167	8	–	–	–	7
	i	–	–	–	–	–	–	–	–	–	7	192	–	–	–	1
	oʊ	–	4	14	1	21	–	10	8	2	–	–	139	–	1	–
	ɔɪ	–	–	–	–	–	–	–	–	–	–	–	–	173	26	1
	ʊ	–	–	–	14	–	–	–	–	–	–	–	–	13	173	–
	u	–	–	–	–	–	–	–	9	–	9	–	–	–	–	182

Table 47: A confusion matrix indicating how the predictions of the multinomial logistic regression applied to GAI's ADS data aligned with the actual identity of vowels in the resampled data.

B Regression coefficients

As in the previous appendix, the following tables provide further detail concerning the results of the multinomial logistic regressions that were presented in 5.4.1. These tables provide values for the estimate and standard error of each coefficient in the regression models across speakers and registers. Further to this, they indicate the z-scores and p-values that were considered when discussing the significance of each acoustic dimension as a predictor of identity. Tables 29–32 provide partial summaries of the significant coefficients of these models for the third formant, log vowel duration, the change in F_1, and the change in F_2 respectively. Recall that /ɛ/ was selected as the reference category and thus no coefficients will be reported for this category. Tables 48–51 detail the full set of coefficients for models that were applied to the IDS productions of ALI, ANN, CIN, and GAI respectively. Tables 52–55 do so for the models that were applied to their ADS productions.

	β	SE	z	p		β	SE	z	p
ɑ (int)	−6.722	0.468	−14.372	< .001	ɑ dur	−0.299	0.184	−1.627	.052
æ (int)	−0.795	0.187	−4.252	< .001	æ dur	−0.092	0.131	−0.703	.241
ʌ (int)	−1.206	0.222	−5.422	< .001	ʌ dur	−0.608	0.158	−3.840	< .001
ɔ (int)	−8.353	0.524	−15.934	< .001	ɔ dur	0.280	0.188	1.491	.068
aʊ (int)	−1.294	0.214	−6.051	< .001	aʊ dur	0.197	0.140	1.408	.080
ɑɪ (int)	−5.750	0.474	−12.127	< .001	ɑɪ dur	−0.438	0.198	−2.209	.014
ɝ (int)	−1.232	0.225	−5.480	< .001	ɝ dur	−0.345	0.148	−2.331	.010
eɪ (int)	−2.114	0.275	−7.676	< .001	eɪ dur	0.389	0.149	2.604	.005
ɪ (int)	−2.555	0.283	−9.041	< .001	ɪ dur	−0.640	0.148	−4.335	< .001
i (int)	−8.715	0.743	−11.722	< .001	i dur	−0.076	0.207	−0.368	.356
oʊ (int)	−4.673	0.385	−12.141	< .001	oʊ dur	0.772	0.174	4.449	< .001
ɔɪ (int)	−13.218	1.326	−9.966	< .001	ɔɪ dur	3.058	0.386	7.925	< .001
ʊ (int)	−2.768	0.274	−10.104	< .001	ʊ dur	0.159	0.173	0.919	.179
u (int)	−7.681	0.568	−13.529	< .001	u dur	0.729	0.197	3.703	< .001
ɑ F_1	5.247	0.359	14.628	< .001	ɑ ΔF_1	0.095	0.146	0.647	.259
æ F_1	2.433	0.249	9.773	< .001	æ ΔF_1	0.198	0.100	1.983	.024
ʌ F_1	3.140	0.301	10.429	< .001	ʌ ΔF_1	−0.293	0.124	−2.365	.009
ɔ F_1	5.393	0.367	14.680	< .001	ɔ ΔF_1	−0.232	0.149	−1.552	.060
aʊ F_1	2.766	0.266	10.399	< .001	aʊ ΔF_1	0.195	0.109	1.791	.037
ɑɪ F_1	6.360	0.385	16.502	< .001	ɑɪ ΔF_1	−0.284	0.145	−1.958	.025
ɝ F_1	−4.206	0.361	−11.663	< .001	ɝ ΔF_1	−0.255	0.125	−2.030	.021
eɪ F_1	1.781	0.284	6.270	< .001	eɪ ΔF_1	−1.112	0.129	−8.601	< .001
ɪ F_1	−4.337	0.381	−11.376	< .001	ɪ ΔF_1	−0.125	0.129	−0.975	.165
i F_1	−5.705	0.574	−9.935	< .001	iΔF_1	−1.015	0.202	−5.014	< .001
oʊ F_1	2.207	0.350	6.298	< .001	oʊ ΔF_1	−0.985	0.150	−6.579	< .001
ɔɪ F_1	−1.466	0.688	−2.131	.017	ɔɪ ΔF_1	−0.506	0.325	−1.555	.060
ʊ F_1	−2.425	0.387	−6.265	< .001	ʊΔF_1	−0.364	0.154	−2.359	.009
u F_1	−10.209	0.584	−17.472	< .001	uΔF_1	−0.621	0.209	−2.974	.001
ɑ F_2	−11.490	0.561	−20.465	< .001	ɑ ΔF_2	−0.196	0.225	−0.871	.192
æ F_2	0.511	0.304	1.685	.046	æ ΔF_2	−0.341	0.159	−2.149	.016
ʌ F_2	−5.152	0.419	−12.305	< .001	ʌ ΔF_2	0.905	0.190	4.777	< .001
ɔ F_2	−12.524	0.579	−21.641	< .001	ɔ ΔF_2	−0.521	0.232	−2.244	.012
aʊ F_2	−1.003	0.337	−2.976	.001	aʊ ΔF_2	−1.401	0.183	−7.651	< .001
ɑɪ F_2	−6.752	0.568	−11.883	< .001	ɑɪ ΔF_2	1.859	0.230	8.078	< .001
ɝ F_2	−2.404	0.314	−7.646	< .001	ɝ ΔF_2	0.649	0.173	3.748	< .001
eɪ F_2	2.709	0.376	7.202	< .001	eɪ ΔF_2	2.043	0.192	10.648	< .001
ɪ F_2	0.604	0.308	1.958	.025	ɪ ΔF_2	0.157	0.163	0.961	.168
i F_2	4.667	0.530	8.813	< .001	i ΔF_2	2.016	0.258	7.799	< .001
oʊ F_2	−10.479	0.530	−19.775	< .001	oʊ ΔF_2	−1.070	0.220	−4.866	< .001
ɔɪ F_2	−13.129	0.857	−15.314	< .001	ɔɪ ΔF_2	0.858	0.282	3.041	.001
ʊ F_2	−8.382	0.466	−18.003	< .001	ʊ ΔF_2	−0.117	0.206	−0.570	.284
u F_2	−5.481	0.439	−12.483	< .001	u ΔF_2	−0.733	0.214	−3.422	< .001
ɑ F_3	−0.075	0.234	−0.321	.374					
æ F_3	0.019	0.164	0.115	.454					
ʌ F_3	−0.050	0.195	−0.257	.399					
ɔ F_3	−0.588	0.242	−2.424	.008					
aʊ F_3	−0.326	0.183	−1.784	.037					
ɑɪ F_3	0.295	0.247	1.194	.116					
ɝ F_3	0.228	0.164	1.389	.082					
eɪ F_3	−0.341	0.186	−1.837	.033					
ɪ F_3	0.077	0.166	0.465	.321					
i F_3	0.795	0.256	3.106	.001					
oʊ F_3	0.098	0.234	0.418	.338					
ɔɪ F_3	0.504	0.366	1.376	.084					
ʊ F_3	0.435	0.214	2.032	.021					
u F_3	0.486	0.257	1.886	.030					

Table 48: The outputs of a multinomial logistic regression for speaker ALI's IDS productions

	β	SE	z	p		β	SE	z	p
ɑ (int)	−4.076	0.400	−10.200	< .001	ɑ dur	0.177	0.182	0.969	.166
æ (int)	−0.282	0.213	−1.325	.093	æ dur	0.535	0.131	4.067	< .001
ʌ (int)	−2.166	0.321	−6.750	< .001	ʌ dur	−0.027	0.174	−0.158	.437
ɔ (int)	−10.471	0.774	−13.535	< .001	ɔ dur	0.805	0.204	3.948	< .001
aʊ (int)	−2.034	0.309	−6.587	< .001	aʊ dur	1.371	0.167	8.188	< .001
ɑɪ (int)	−3.434	0.409	−8.396	< .001	ɑɪ dur	0.463	0.182	2.537	.006
ɝ (int)	−0.927	0.264	−3.510	< .001	ɝ dur	1.143	0.154	7.408	< .001
eɪ (int)	−2.237	0.346	−6.464	< .001	eɪ dur	1.129	0.171	6.606	< .001
ɪ (int)	−1.933	0.296	−6.520	< .001	ɪ dur	0.139	0.143	0.970	.166
i (int)	−11.259	1.063	−10.592	< .001	i dur	0.314	0.229	1.372	.085
oʊ (int)	−2.585	0.333	−7.767	< .001	oʊ dur	0.945	0.172	5.503	< .001
ɔɪ (int)	−20.095	2.690	−7.471	< .001	ɔɪ dur	−0.290	0.418	−0.694	.244
ʊ (int)	−5.228	0.491	−10.643	< .001	ʊ dur	−0.631	0.217	−2.905	.002
u (int)	−8.192	0.720	−11.371	< .001	u dur	0.942	0.202	4.658	< .001
ɑ F_1	1.841	0.356	5.173	< .001	ɑ ΔF_1	−0.498	0.231	−2.156	.016
æ F_1	2.680	0.288	9.317	< .001	æ ΔF_1	−0.430	0.177	−2.433	.007
ʌ F_1	2.183	0.342	6.392	< .001	ʌ ΔF_1	−0.883	0.219	−4.035	< .001
ɔ F_1	1.556	0.406	3.834	< .001	ɔ ΔF_1	−0.186	0.271	−0.687	.246
aʊ F_1	3.187	0.343	9.292	< .001	aʊ ΔF_1	−1.125	0.208	−5.402	< .001
ɑɪ F_1	3.883	0.385	10.079	< .001	ɑɪ ΔF_1	−1.452	0.241	−6.014	< .001
ɝ F_1	−5.159	0.417	−12.380	< .001	ɝ ΔF_1	−2.132	0.258	−8.255	< .001
eɪ F_1	1.059	0.412	2.570	.005	eɪ ΔF_1	−3.090	0.300	−10.294	< .001
ɪ F_1	−4.266	0.401	−10.651	< .001	ɪ ΔF_1	−1.078	0.231	−4.661	< .001
i F_1	−6.977	0.712	−9.798	< .001	i ΔF_1	−3.456	0.464	−7.453	< .001
oʊ F_1	−1.643	0.374	−4.395	< .001	oʊ ΔF_1	−2.294	0.247	−9.296	< .001
ɔɪ F_1	−8.696	1.331	−6.535	< .001	ɔɪ ΔF_1	−1.781	0.594	−2.999	.001
ʊ F_1	−4.555	0.458	−9.948	< .001	ʊ ΔF_1	−2.302	0.288	−8.002	< .001
u F_1	−11.269	0.700	−16.094	< .001	u ΔF_1	−2.722	0.349	−7.808	< .001
ɑ F_2	−9.731	0.515	−18.883	< .001	ɑ ΔF_2	0.986	0.247	3.990	< .001
æ F_2	0.562	0.331	1.698	.045	æ ΔF_2	0.246	0.178	1.385	.083
ʌ F_2	−7.530	0.486	−15.495	< .001	ʌ ΔF_2	0.351	0.234	1.502	.067
ɔ F_2	−14.666	0.659	−22.269	< .001	ɔ ΔF_2	−0.322	0.291	−1.106	.134
aʊ F_2	−4.975	0.452	−10.995	< .001	aʊ ΔF_2	−1.199	0.233	−5.146	< .001
ɑɪ F_2	−4.391	0.540	−8.125	< .001	ɑɪ ΔF_2	2.697	0.284	9.510	< .001
ɝ F_2	−3.720	0.363	−10.261	< .001	ɝ ΔF_2	0.482	0.196	2.456	.007
eɪ F_2	2.447	0.465	5.258	< .001	eɪ ΔF_2	2.224	0.266	8.357	< .001
ɪ F_2	0.012	0.324	0.037	.485	ɪ ΔF_2	−0.048	0.163	−0.292	.385
i F_2	5.320	0.702	7.578	< .001	i ΔF_2	0.751	0.289	2.600	.005
oʊ F_2	−8.691	0.488	−17.799	< .001	oʊ ΔF_2	−1.346	0.242	−5.568	< .001
ɔɪ F_2	−12.318	1.117	−11.029	< .001	ɔɪ ΔF_2	5.417	0.695	7.791	< .001
ʊ F_2	−9.204	0.511	−18.006	< .001	ʊ ΔF_2	1.961	0.285	6.876	< .001
u F_2	−5.811	0.439	−13.251	< .001	u ΔF_2	−1.839	0.253	−7.275	< .001
ɑ F_3	−0.337	0.254	−1.323	.093					
æ F_3	0.126	0.164	0.765	.222					
ʌ F_3	−0.219	0.245	−0.892	.186					
ɔ F_3	−0.707	0.285	−2.482	.007					
aʊ F_3	0.339	0.238	1.423	.077					
ɑɪ F_3	−0.498	0.268	−1.859	.032					
ɝ F_3	0.454	0.157	2.881	.002					
eɪ F_3	0.097	0.185	0.526	.299					
ɪ F_3	0.185	0.144	1.287	.099					
i F_3	0.322	0.222	1.451	.073					
oʊ F_3	−0.020	0.238	−0.084	.466					
ɔɪ F_3	2.619	0.589	4.443	< .001					
ʊ F_3	0.376	0.250	1.504	.066					
u F_3	−0.270	0.245	−1.101	.136					

Table 49: The outputs of a multinomial logistic regression for speaker ANN's IDS productions

	β	SE	z	p		β	SE	z	p
ɑ (int)	−4.108	0.439	−9.366	< .001	ɑ dur	−0.060	0.200	−0.302	.381
æ (int)	−1.642	0.323	−5.085	< .001	æ dur	−0.006	0.161	−0.037	.485
ʌ (int)	−0.489	0.292	−1.672	.047	ʌ dur	−0.571	0.188	−3.045	.001
ɔ (int)	−9.430	0.698	−13.504	< .001	ɔ dur	0.052	0.221	0.237	.407
aʊ (int)	−2.764	0.396	−6.972	< .001	aʊ dur	0.336	0.184	1.828	.034
ɑɪ (int)	−5.868	0.630	−9.316	< .001	ɑɪ dur	0.246	0.220	1.116	.132
ɝ (int)	1.109	0.229	4.849	< .001	ɝ dur	0.173	0.162	1.064	.144
eɪ (int)	−3.180	0.418	−7.614	< .001	eɪ dur	0.781	0.175	4.471	< .001
ɪ (int)	−1.427	0.310	−4.598	< .001	ɪ dur	−0.255	0.159	−1.599	.055
i (int)	−12.323	1.129	−10.917	< .001	i dur	0.168	0.216	0.777	.219
oʊ (int)	−1.306	0.337	−3.873	< .001	oʊ dur	0.884	0.193	4.576	< .001
ɔɪ (int)	−23.975	3.118	−7.690	< .001	ɔɪ dur	−0.550	0.464	−1.186	.118
ʊ (int)	−4.202	0.472	−8.905	< .001	ʊ dur	−0.206	0.218	−0.948	.172
u (int)	−8.395	0.897	−9.358	< .001	u dur	0.399	0.279	1.431	.076
ɑ F_1	4.161	0.406	10.260	< .001	ɑ ΔF_1	0.076	0.169	0.448	.327
æ F_1	4.263	0.359	11.876	< .001	æ ΔF_1	−0.225	0.136	−1.658	.049
ʌ F_1	2.491	0.369	6.760	< .001	ʌ ΔF_1	−0.547	0.155	−3.538	< .001
ɔ F_1	2.666	0.449	5.939	< .001	ɔ ΔF_1	0.562	0.195	2.876	.002
aʊ F_1	4.573	0.398	11.484	< .001	aʊ ΔF_1	−0.086	0.150	−0.574	.283
ɑɪ F_1	6.261	0.496	12.614	< .001	ɑɪ ΔF_1	−1.115	0.185	−6.022	< .001
ɝ F_1	−3.328	0.331	−10.042	< .001	ɝ ΔF_1	−1.362	0.163	−8.338	< .001
eɪ F_1	−1.604	0.345	−4.653	< .001	eɪ ΔF_1	−2.376	0.209	−11.393	< .001
ɪ F_1	−4.550	0.360	−12.640	< .001	ɪ ΔF_1	−0.641	0.177	−3.620	< .001
i F_1	−7.066	0.607	−11.633	< .001	i ΔF_1	−1.446	0.304	−4.753	< .001
oʊ F_1	0.533	0.394	1.353	.088	oʊ ΔF_1	−0.695	0.164	−4.224	< .001
ɔɪ F_1	−2.296	0.704	−3.260	.001	ɔɪ ΔF_1	−1.600	0.730	−2.192	.014
ʊ F_1	−0.987	0.440	−2.245	.012	ʊ ΔF_1	−0.573	0.189	−3.030	.001
u F_1	−9.688	0.822	−11.793	< .001	u ΔF_1	−1.547	0.344	−4.491	< .001
ɑ F_2	−12.315	0.633	−19.468	< .001	ɑ ΔF_2	0.613	0.275	2.228	.013
æ F_2	−2.105	0.448	−4.693	< .001	æ ΔF_2	−0.049	0.193	−0.256	.399
ʌ F_2	−9.003	0.564	−15.950	< .001	ʌ ΔF_2	0.718	0.239	3.007	.001
ɔ F_2	−17.763	0.742	−23.941	< .001	ɔ ΔF_2	−0.453	0.306	−1.484	.069
aʊ F_2	−7.355	0.560	−13.144	< .001	aʊ ΔF_2	−2.254	0.265	−8.495	< .001
ɑɪ F_2	−7.428	0.667	−11.142	< .001	ɑɪ ΔF_2	3.240	0.340	9.517	< .001
ɝ F_2	−3.303	0.419	−7.887	< .001	ɝ ΔF_2	0.338	0.194	1.745	.040
eɪ F_2	3.752	0.494	7.595	< .001	eɪ ΔF_2	1.769	0.226	7.836	< .001
ɪ F_2	0.793	0.397	2.000	.023	ɪ ΔF_2	0.316	0.194	1.627	.052
i F_2	7.674	0.751	10.212	< .001	i ΔF_2	2.417	0.344	7.021	< .001
oʊ F_2	−10.868	0.608	−17.889	< .001	oʊ ΔF_2	−1.011	0.260	−3.882	< .001
ɔɪ F_2	−17.368	1.433	−12.121	< .001	ɔɪ ΔF_2	7.848	1.137	6.905	< .001
ʊ F_2	−13.524	0.676	−20.015	< .001	ʊ ΔF_2	0.547	0.279	1.965	.025
u F_2	−9.155	0.718	−12.748	< .001	u ΔF_2	−1.288	0.354	−3.644	< .001
ɑ F_3	0.017	0.257	0.065	.474					
æ F_3	0.328	0.215	1.528	.063					
ʌ F_3	0.508	0.235	2.167	.015					
ɔ F_3	0.618	0.277	2.234	.013					
aʊ F_3	0.100	0.254	0.394	.347					
ɑɪ F_3	0.255	0.306	0.832	.203					
ɝ F_3	0.131	0.183	0.714	.237					
eɪ F_3	−0.382	0.215	−1.779	.038					
ɪ F_3	0.052	0.183	0.284	.388					
i F_3	0.087	0.302	0.289	.386					
oʊ F_3	0.349	0.248	1.405	.080					
ɔɪ F_3	1.411	0.683	2.067	.019					
ʊ F_3	0.599	0.267	2.245	.012					
u F_3	0.539	0.373	1.448	.074					

Table 50: The outputs of a multinomial logistic regression for speaker CIN's IDS productions

	β	SE	z	p		β	SE	z	p
ɑ (int)	−5.284	0.455	−11.624	< .001	ɑ dur	0.549	0.190	2.896	.002
æ (int)	−1.960	0.278	−7.062	< .001	æ dur	1.299	0.174	7.464	< .001
ʌ (int)	−1.885	0.288	−6.536	< .001	ʌ dur	−0.092	0.179	−0.511	.305
ɔ (int)	−5.734	0.465	−12.330	< .001	ɔ dur	1.153	0.195	5.906	< .001
aʊ (int)	−2.199	0.301	−7.306	< .001	aʊ dur	1.025	0.179	5.722	< .001
ɑɪ (int)	−5.748	0.546	−10.536	< .001	ɑɪ dur	0.589	0.207	2.841	.002
ɝ (int)	0.134	0.181	0.736	.231	ɝ dur	0.527	0.170	3.107	.001
eɪ (int)	−2.272	0.324	−7.016	< .001	eɪ dur	1.110	0.195	5.691	< .001
ɪ (int)	−2.032	0.290	−7.009	< .001	ɪ dur	−0.217	0.189	−1.145	.126
i (int)	−13.746	1.470	−9.352	< .001	i dur	0.743	0.298	2.493	.006
oʊ (int)	−4.706	0.418	−11.260	< .001	oʊ dur	1.110	0.188	5.896	< .001
ɔɪ (int)	−14.480	1.231	−11.762	< .001	ɔɪ dur	3.067	0.303	10.132	< .001
ʊ (int)	−4.016	0.385	−10.441	< .001	ʊ dur	0.537	0.192	2.800	.003
u (int)	−10.347	0.914	−11.320	< .001	u dur	0.644	0.262	2.459	.007
ɑ F_1	3.359	0.383	8.764	< .001	ɑ ΔF_1	0.320	0.187	1.709	.044
æ F_1	3.746	0.333	11.234	< .001	æ ΔF_1	−0.155	0.153	−1.013	.155
ʌ F_1	2.003	0.340	5.889	< .001	ʌ ΔF_1	−0.747	0.168	−4.455	< .001
ɔ F_1	−1.048	0.421	−2.492	.006	ɔ ΔF_1	−0.016	0.210	−0.077	.469
aʊ F_1	3.403	0.351	9.701	< .001	aʊ ΔF_1	−0.866	0.164	−5.275	< .001
ɑɪ F_1	5.088	0.442	11.522	< .001	ɑɪ ΔF_1	−1.330	0.190	−7.007	< .001
ɝ F_1	−3.668	0.358	−10.247	< .001	ɝ ΔF_1	−0.918	0.172	−5.332	< .001
eɪ F_1	0.141	0.372	0.381	.352	eɪ ΔF_1	−1.826	0.197	−9.275	< .001
ɪ F_1	−5.193	0.424	−12.237	< .001	ɪ ΔF_1	−0.594	0.190	−3.131	.001
i F_1	−11.630	1.019	−11.410	< .001	i ΔF_1	−1.657	0.343	−4.825	< .001
oʊ F_1	−0.242	0.392	−0.617	.269	oʊ ΔF_1	−1.962	0.197	−9.946	< .001
ɔɪ F_1	−1.208	0.613	−1.972	.024	ɔɪ ΔF_1	−0.499	0.320	−1.561	.059
ʊ F_1	−1.594	0.400	−3.987	< .001	ʊ ΔF_1	−0.274	0.200	−1.372	.085
u F_1	−12.979	0.810	−16.023	< .001	u ΔF_1	−0.955	0.319	−2.996	.001
ɑ F_2	−9.933	0.552	−17.993	< .001	ɑ ΔF_2	0.225	0.239	0.943	.173
æ F_2	0.801	0.390	2.057	.020	æ ΔF_2	0.860	0.195	4.419	< .001
ʌ F_2	−6.398	0.470	−13.623	< .001	ʌ ΔF_2	1.018	0.213	4.784	< .001
ɔ F_2	−11.080	0.566	−19.567	< .001	ɔ ΔF_2	−0.471	0.247	−1.907	.028
aʊ F_2	−3.647	0.440	−8.295	< .001	aʊ ΔF_2	−1.243	0.221	−5.614	< .001
ɑɪ F_2	−4.683	0.574	−8.165	< .001	ɑɪ ΔF_2	3.390	0.293	11.577	< .001
ɝ F_2	−3.152	0.373	−8.447	< .001	ɝ ΔF_2	0.488	0.183	2.662	.004
eɪ F_2	3.804	0.465	8.174	< .001	eɪ ΔF_2	1.896	0.207	9.156	< .001
ɪ F_2	0.713	0.380	1.879	.030	ɪ ΔF_2	0.723	0.175	4.137	< .001
i F_2	5.337	0.810	6.591	< .001	i ΔF_2	1.815	0.319	5.695	< .001
oʊ F_2	−10.172	0.540	−18.821	< .001	oʊ ΔF_2	−0.369	0.241	−1.528	.063
ɔɪ F_2	−14.428	0.833	−17.323	< .001	ɔɪ ΔF_2	0.897	0.311	2.882	.002
ʊ F_2	−10.023	0.536	−18.713	< .001	ʊ ΔF_2	0.417	0.237	1.763	.039
u F_2	−6.372	0.501	−12.711	< .001	u ΔF_2	−1.856	0.279	−6.652	< .001
ɑ F_3	−0.721	0.213	−3.382	< .001					
æ F_3	−0.467	0.215	−2.177	.015					
ʌ F_3	−0.349	0.203	−1.723	.042					
ɔ F_3	−0.713	0.220	−3.238	.001					
aʊ F_3	−0.663	0.217	−3.060	.001					
ɑɪ F_3	−0.657	0.241	−2.725	.003					
ɝ F_3	−0.659	0.185	−3.569	< .001					
eɪ F_3	−0.555	0.228	−2.439	.007					
ɪ F_3	0.750	0.215	3.490	< .001					
i F_3	1.324	0.385	3.440	< .001					
oʊ F_3	−0.627	0.216	−2.895	.002					
ɔɪ F_3	−0.346	0.319	−1.084	.139					
ʊ F_3	−0.071	0.215	−0.329	.371					
u F_3	0.826	0.322	2.569	.005					

Table 51: The outputs of a multinomial logistic regression for speaker GAI's IDS productions

	β	SE	z	p		β	SE	z	p
ɑ (int)	−16.350	1.122	−14.567	< .001	ɑ dur	0.827	0.257	3.213	.001
æ (int)	−2.085	0.302	−6.893	< .001	æ dur	0.429	0.168	2.561	.005
ʌ (int)	−1.166	0.264	−4.426	< .001	ʌ dur	0.632	0.171	3.686	< .001
ɔ (int)	−20.349	1.297	−15.694	< .001	ɔ dur	0.733	0.268	2.736	.003
aʊ (int)	−5.083	0.462	−10.996	< .001	aʊ dur	1.518	0.207	7.343	< .001
ɑɪ (int)	−9.042	0.754	−11.988	< .001	ɑɪ dur	−0.213	0.238	−0.892	.186
ɝ (int)	−0.204	0.212	−0.961	.168	ɝ dur	0.853	0.170	5.016	< .001
eɪ (int)	−1.669	0.319	−5.224	< .001	eɪ dur	1.027	0.207	4.973	< .001
ɪ (int)	−2.374	0.339	−7.010	< .001	ɪ dur	−0.191	0.179	−1.070	.142
i (int)	−9.477	0.798	−11.870	< .001	i dur	−0.189	0.228	−0.829	.204
oʊ (int)	−3.045	0.371	−8.211	< .001	oʊ dur	0.940	0.187	5.015	< .001
ɔɪ (int)	−28.306	5.529	−5.120	< .001	ɔɪ dur	2.351	1.031	2.280	.011
ʊ (int)	−4.330	0.426	−10.173	< .001	ʊ dur	1.033	0.216	4.777	< .001
u (int)	−9.182	0.750	−12.241	< .001	u dur	1.736	0.234	7.410	< .001
ɑ F_1	8.532	0.582	14.671	< .001	ɑ ΔF_1	0.714	0.284	2.518	.006
æ F_1	4.923	0.408	12.076	< .001	æ ΔF_1	−0.385	0.168	−2.301	.011
ʌ F_1	3.385	0.383	8.832	< .001	ʌ ΔF_1	−0.617	0.184	−3.348	< .001
ɔ F_1	8.415	0.601	13.999	< .001	ɔ ΔF_1	1.187	0.311	3.817	< .001
aʊ F_1	7.027	0.478	14.695	< .001	aʊ ΔF_1	−0.662	0.193	−3.432	< .001
ɑɪ F_1	8.828	0.561	15.735	< .001	ɑɪ ΔF_1	−0.213	0.245	−0.868	.193
ɝ F_1	−3.585	0.387	−9.268	< .001	ɝ ΔF_1	−1.496	0.192	−7.800	< .001
eɪ F_1	2.968	0.436	6.815	< .001	eɪ ΔF_1	−1.539	0.200	−7.709	< .001
ɪ F_1	−5.259	0.475	−11.076	< .001	ɪ ΔF_1	−0.337	0.199	−1.695	.045
i F_1	−9.919	0.736	−13.474	< .001	i ΔF_1	−1.314	0.266	−4.949	< .001
oʊ F_1	2.476	0.416	5.960	< .001	oʊ ΔF_1	−1.657	0.205	−8.093	< .001
ɔɪ F_1	1.375	1.953	0.704	.241	ɔɪ ΔF_1	2.041	0.933	2.187	.014
ʊ F_1	−4.098	0.527	−7.783	< .001	ʊ ΔF_1	−1.414	0.241	−5.869	< .001
u F_1	−13.744	0.762	−18.040	< .001	u ΔF_1	−1.289	0.268	−4.813	< .001
ɑ F_2	−16.815	0.850	−19.774	< .001	ɑ ΔF_2	2.284	0.357	6.396	< .001
æ F_2	1.651	0.391	4.224	< .001	æ ΔF_2	−0.470	0.231	−2.034	.021
ʌ F_2	−5.427	0.437	−12.411	< .001	ʌ ΔF_2	1.351	0.250	5.403	< .001
ɔ F_2	−20.613	0.930	−22.176	< .001	ɔ ΔF_2	0.293	0.383	0.766	.222
aʊ F_2	0.416	0.514	0.810	.209	aʊ ΔF_2	−1.586	0.296	−5.356	< .001
ɑɪ F_2	−7.802	0.716	−10.899	< .001	ɑɪ ΔF_2	3.378	0.336	10.053	< .001
ɝ F_2	−4.512	0.374	−12.055	< .001	ɝ ΔF_2	0.860	0.237	3.634	< .001
eɪ F_2	2.058	0.452	4.553	< .001	eɪ ΔF_2	3.388	0.281	12.077	< .001
ɪ F_2	0.679	0.333	2.037	.021	ɪ ΔF_2	1.572	0.231	6.820	< .001
i F_2	2.862	0.467	6.127	< .001	i ΔF_2	1.207	0.259	4.669	< .001
oʊ F_2	−7.919	0.524	−15.113	< .001	oʊ ΔF_2	−1.388	0.293	−4.731	< .001
ɔɪ F_2	−15.660	2.189	−7.154	< .001	ɔɪ ΔF_2	10.540	2.187	4.819	< .001
ʊ F_2	−8.228	0.499	−16.498	< .001	ʊ ΔF_2	2.564	0.331	7.755	< .001
u F_2	−2.960	0.388	−7.636	< .001	u ΔF_2	0.603	0.275	2.196	.014
ɑ F_3	−1.813	0.261	−6.957	< .001					
æ F_3	−0.623	0.150	−4.148	< .001					
ʌ F_3	−0.483	0.167	−2.893	.002					
ɔ F_3	−1.793	0.269	−6.671	< .001					
aʊ F_3	−0.607	0.182	−3.340	< .001					
ɑɪ F_3	−1.171	0.238	−4.929	< .001					
ɝ F_3	−0.164	0.145	−1.130	.129					
eɪ F_3	−0.666	0.187	−3.566	< .001					
ɪ F_3	−0.546	0.151	−3.615	< .001					
i F_3	−0.434	0.199	−2.186	.014					
oʊ F_3	−0.590	0.185	−3.182	.001					
ɔɪ F_3	1.357	0.935	1.452	.073					
ʊ F_3	−0.187	0.194	−0.964	.168					
u F_3	−0.430	0.198	−2.169	.015					

Table 52: The outputs of a multinomial logistic regression for speaker ALI's ADS productions

	β	SE	z	p		β	SE	z	p
ɑ (int)	−7.308	0.520	−14.041	< .001	ɑ dur	0.241	0.176	1.365	.086
æ (int)	−1.232	0.220	−5.612	< .001	æ dur	0.589	0.148	3.981	< .001
ʌ (int)	−2.610	0.300	−8.706	< .001	ʌ dur	0.215	0.158	1.356	.088
ɔ (int)	−11.897	0.737	−16.151	< .001	ɔ dur	−0.320	0.194	−1.645	.050
aʊ (int)	−2.984	0.311	−9.587	< .001	aʊ dur	0.513	0.161	3.196	.001
ɑɪ (int)	−5.740	0.495	−11.589	< .001	ɑɪ dur	−0.011	0.205	−0.055	.478
ɝ (int)	−0.414	0.197	−2.105	.018	ɝ dur	1.074	0.164	6.532	< .001
eɪ (int)	−0.868	0.236	−3.686	< .001	eɪ dur	0.784	0.171	4.570	< .001
ɪ (int)	−3.416	0.371	−9.211	< .001	ɪ dur	−0.271	0.207	−1.306	.096
i (int)	−8.380	0.658	−12.727	< .001	i dur	0.370	0.256	1.444	.074
oʊ (int)	−1.793	0.264	−6.794	< .001	oʊ dur	0.611	0.159	3.845	< .001
ɔɪ (int)	−12.302	1.430	−8.602	< .001	ɔɪ dur	0.042	0.359	0.116	.454
ʊ (int)	−9.876	0.888	−11.121	< .001	ʊ dur	−0.623	0.251	−2.479	.007
u (int)	−12.208	0.998	−12.235	< .001	u dur	1.366	0.280	4.887	< .001
ɑ F_1	4.167	0.414	10.057	< .001	ɑ ΔF_1	0.739	0.205	3.608	< .001
æ F_1	2.904	0.339	8.555	< .001	æ ΔF_1	0.301	0.154	1.958	.025
ʌ F_1	1.842	0.370	4.978	< .001	ʌ ΔF_1	−0.209	0.186	−1.122	.131
ɔ F_1	3.926	0.443	8.870	< .001	ɔ ΔF_1	0.823	0.219	3.761	< .001
aʊ F_1	3.128	0.374	8.370	< .001	aʊ ΔF_1	−0.197	0.184	−1.071	.142
ɑɪ F_1	5.966	0.454	13.133	< .001	ɑɪ ΔF_1	−0.813	0.230	−3.534	< .001
ɝ F_1	−6.925	0.504	−13.729	< .001	ɝ ΔF_1	−1.038	0.209	−4.965	< .001
eɪ F_1	−4.705	0.537	−8.769	< .001	eɪ ΔF_1	−1.458	0.240	−6.079	< .001
ɪ F_1	−10.313	0.665	−15.497	< .001	ɪ ΔF_1	−0.530	0.283	−1.873	.031
i F_1	−13.054	0.827	−15.793	< .001	i ΔF_1	−0.997	0.353	−2.826	.002
oʊ F_1	−1.317	0.383	−3.436	< .001	oʊ ΔF_1	−0.605	0.188	−3.217	.001
ɔɪ F_1	−11.422	0.955	−11.964	< .001	ɔɪ ΔF_1	−2.102	0.417	−5.042	< .001
ʊ F_1	−2.871	0.574	−5.005	< .001	ʊ ΔF_1	−0.412	0.241	−1.706	.044
u F_1	−20.027	1.038	−19.298	< .001	u ΔF_1	−2.429	0.381	−6.370	< .001
ɑ F_2	−10.345	0.555	−18.636	< .001	ɑ ΔF_2	0.310	0.217	1.430	.076
æ F_2	0.413	0.353	1.170	.121	æ ΔF_2	−0.193	0.181	−1.064	.144
ʌ F_2	−7.076	0.460	−15.369	< .001	ʌ ΔF_2	−0.079	0.200	−0.394	.347
ɔ F_2	−14.405	0.644	−22.378	< .001	ɔ ΔF_2	−0.163	0.230	−0.709	.239
aʊ F_2	−6.153	0.462	−13.311	< .001	aʊ ΔF_2	−0.244	0.202	−1.210	.113
ɑɪ F_2	−3.655	0.571	−6.402	< .001	ɑɪ ΔF_2	1.916	0.256	7.490	< .001
ɝ F_2	−2.935	0.371	−7.905	< .001	ɝ ΔF_2	1.145	0.216	5.309	< .001
eɪ F_2	1.446	0.417	3.471	< .001	eɪ ΔF_2	2.528	0.255	9.900	< .001
ɪ F_2	1.082	0.439	2.462	.007	ɪ ΔF_2	1.720	0.269	6.395	< .001
i F_2	3.964	0.575	6.893	< .001	i ΔF_2	1.729	0.306	5.649	< .001
oʊ F_2	−6.868	0.441	−15.590	< .001	oʊ ΔF_2	−0.707	0.198	−3.578	< .001
ɔɪ F_2	−8.391	0.978	−8.579	< .001	ɔɪ ΔF_2	5.284	0.535	9.870	< .001
ʊ F_2	−13.031	0.735	−17.739	< .001	ʊ ΔF_2	1.974	0.289	6.838	< .001
u F_2	−2.156	0.489	−4.410	< .001	u ΔF_2	0.639	0.295	2.167	.015
ɑ F_3	−0.005	0.185	−0.026	.490					
æ F_3	0.266	0.156	1.709	.044					
ʌ F_3	0.615	0.170	3.619	< .001					
ɔ F_3	−0.502	0.200	−2.508	.006					
aʊ F_3	0.682	0.172	3.955	< .001					
ɑɪ F_3	0.311	0.222	1.398	.081					
ɝ F_3	−0.316	0.160	−1.969	.024					
eɪ F_3	−0.356	0.178	−2.005	.022					
ɪ F_3	−0.557	0.196	−2.847	.002					
i F_3	−0.173	0.231	−0.750	.227					
oʊ F_3	0.221	0.166	1.330	.092					
ɔɪ F_3	1.902	0.488	3.900	< .001					
ʊ F_3	1.339	0.263	5.091	< .001					
u F_3	−0.562	0.248	−2.263	.012					

Table 53: The outputs of a multinomial logistic regression for speaker ANN's ADS productions

	β	SE	z	p		β	SE	z	p
ɑ (int)	−7.420	0.529	−14.037	< .001	ɑ dur	0.924	0.219	4.217	< .001
æ (int)	−3.791	0.378	−10.033	< .001	æ dur	0.498	0.218	2.281	.011
ʌ (int)	−2.633	0.306	−8.603	< .001	ʌ dur	−0.061	0.199	−0.306	.380
ɔ (int)	−11.058	0.690	−16.015	< .001	ɔ dur	0.868	0.231	3.765	< .001
aʊ (int)	−5.012	0.431	−11.618	< .001	aʊ dur	1.287	0.225	5.731	< .001
ɑɪ (int)	−6.840	0.613	−11.160	< .001	ɑɪ dur	0.087	0.235	0.370	.356
ɝ (int)	−2.524	0.299	−8.434	< .001	ɝ dur	1.022	0.202	5.060	< .001
eɪ (int)	−1.219	0.235	−5.180	< .001	eɪ dur	0.909	0.206	4.421	< .001
ɪ (int)	−2.084	0.277	−7.522	< .001	ɪ dur	−0.301	0.191	−1.579	.057
i (int)	−12.667	1.220	−10.379	< .001	i dur	−0.279	0.315	−0.886	.188
oʊ (int)	−4.538	0.390	−11.636	< .001	oʊ dur	0.988	0.216	4.579	< .001
ɔɪ (int)	−28.196	2.771	−10.177	< .001	ɔɪ dur	3.432	0.441	7.781	< .001
ʊ (int)	−9.730	0.759	−12.814	< .001	ʊ dur	0.694	0.254	2.733	.003
u (int)	−12.467	0.950	−13.119	< .001	u dur	1.039	0.259	4.016	< .001
ɑ F_1	5.718	0.479	11.925	< .001	ɑ ΔF_1	0.785	0.258	3.043	.001
æ F_1	6.631	0.488	13.579	< .001	æ ΔF_1	−0.164	0.235	−0.697	.243
ʌ F_1	2.683	0.389	6.906	< .001	ʌ ΔF_1	−0.295	0.226	−1.307	.096
ɔ F_1	5.029	0.510	9.863	< .001	ɔ ΔF_1	1.069	0.290	3.681	< .001
aʊ F_1	6.310	0.487	12.946	< .001	aʊ ΔF_1	−0.708	0.243	−2.920	.002
ɑɪ F_1	7.345	0.562	13.065	< .001	ɑɪ ΔF_1	−1.165	0.260	−4.477	< .001
ɝ F_1	−7.456	0.527	−14.155	< .001	ɝ ΔF_1	−1.326	0.272	−4.874	< .001
eɪ F_1	−0.015	0.398	−0.037	.485	eɪ ΔF_1	−2.486	0.266	−9.328	< .001
ɪ F_1	−5.543	0.480	−11.549	< .001	ɪ ΔF_1	−0.754	0.232	−3.249	.001
i F_1	−13.312	1.000	−13.306	< .001	i ΔF_1	−1.856	0.474	−3.912	< .001
oʊ F_1	−0.090	0.448	−0.202	.420	oʊ ΔF_1	−1.745	0.262	−6.672	< .001
ɔɪ F_1	−3.940	0.905	−4.352	< .001	ɔɪ ΔF_1	−5.757	0.608	−9.463	< .001
ʊ F_1	−4.248	0.639	−6.645	< .001	ʊ ΔF_1	−0.768	0.332	−2.310	.010
u F_1	−15.963	0.895	−17.825	< .001	u ΔF_1	−0.121	0.445	−0.272	.393
ɑ F_2	−11.338	0.679	−16.697	< .001	ɑ ΔF_2	−0.615	0.298	−2.062	.020
æ F_2	0.563	0.600	0.938	.174	æ ΔF_2	−0.870	0.297	−2.929	.002
ʌ F_2	−7.431	0.581	−12.799	< .001	ʌ ΔF_2	−0.297	0.268	−1.106	.134
ɔ F_2	−15.522	0.748	−20.761	< .001	ɔ ΔF_2	−1.364	0.325	−4.193	< .001
aʊ F_2	−4.319	0.650	−6.643	< .001	aʊ ΔF_2	−2.377	0.316	−7.532	< .001
ɑɪ F_2	−5.389	0.763	−7.060	< .001	ɑɪ ΔF_2	1.974	0.337	5.860	< .001
ɝ F_2	−3.626	0.474	−7.653	< .001	ɝ ΔF_2	0.522	0.262	1.993	.023
eɪ F_2	4.360	0.544	8.016	< .001	eɪ ΔF_2	2.733	0.283	9.651	< .001
ɪ F_2	1.890	0.426	4.441	< .001	ɪ ΔF_2	1.292	0.237	5.442	< .001
i F_2	7.251	0.874	8.297	< .001	i ΔF_2	2.305	0.447	5.158	< .001
oʊ F_2	−10.461	0.643	−16.277	< .001	oʊ ΔF_2	−1.329	0.308	−4.318	< .001
ɔɪ F_2	−20.721	1.693	−12.241	< .001	ɔɪ ΔF_2	5.438	0.696	7.812	< .001
ʊ F_2	−12.879	0.781	−16.491	< .001	ʊ ΔF_2	0.418	0.363	1.151	.125
u F_2	−4.134	0.586	−7.048	< .001	u ΔF_2	−2.321	0.396	−5.854	< .001
ɑ F_3	−0.790	0.267	−2.960	.002					
æ F_3	−0.281	0.258	−1.089	.138					
ʌ F_3	−0.546	0.217	−2.522	.006					
ɔ F_3	−0.616	0.288	−2.139	.016					
aʊ F_3	−0.385	0.275	−1.400	.081					
ɑɪ F_3	−0.806	0.313	−2.576	.005					
ɝ F_3	−0.145	0.178	−0.811	.209					
eɪ F_3	−0.735	0.213	−3.458	< .001					
ɪ F_3	−0.160	0.175	−0.915	.180					
i F_3	0.037	0.324	0.113	.455					
oʊ F_3	−0.263	0.233	−1.130	.129					
ɔɪ F_3	−1.562	0.559	−2.794	.003					
ʊ F_3	−0.694	0.287	−2.423	.008					
u F_3	−0.149	0.246	−0.603	.273					

Table 54: The outputs of a multinomial logistic regression for speaker CIN's ADS productions

	β	SE	z	p		β	SE	z	p
ɑ (int)	−13.509	0.868	−15.555	< .001	ɑ dur	0.398	0.224	1.780	.038
æ (int)	−6.932	0.580	−11.956	< .001	æ dur	0.822	0.175	4.697	< .001
ʌ (int)	−4.772	0.492	−9.692	< .001	ʌ dur	0.258	0.183	1.408	.080
ɔ (int)	−16.295	1.096	−14.864	< .001	ɔ dur	−1.288	0.278	−4.637	< .001
aʊ (int)	−5.305	0.508	−10.435	< .001	aʊ dur	1.276	0.187	6.824	< .001
ɑɪ (int)	−11.532	0.847	−13.614	< .001	ɑɪ dur	0.545	0.214	2.543	.005
ɝ (int)	0.923	0.293	3.147	.001	ɝ dur	0.512	0.218	2.353	.009
eɪ (int)	−1.441	0.472	−3.053	.001	eɪ dur	1.271	0.240	5.289	< .001
ɪ (int)	−1.106	0.421	−2.625	.004	ɪ dur	−0.831	0.304	−2.731	.003
i (int)	−15.586	2.273	−6.857	< .001	i dur	−0.783	0.478	−1.640	.050
oʊ (int)	−1.408	0.370	−3.803	< .001	oʊ dur	1.047	0.189	5.553	< .001
ɔɪ (int)	−24.965	2.418	−10.325	< .001	ɔɪ dur	0.800	0.484	1.655	.049
ʊ (int)	−14.809	1.758	−8.424	< .001	ʊ dur	−0.433	0.394	−1.100	.136
u (int)	−4.504	0.715	−6.303	< .001	u dur	1.327	0.320	4.148	< .001
ɑ F_1	8.072	0.636	12.700	< .001	ɑ ΔF_1	0.960	0.223	4.300	< .001
æ F_1	8.424	0.632	13.334	< .001	æ ΔF_1	−0.171	0.176	−0.973	.165
ʌ F_1	3.205	0.515	6.224	< .001	ʌ ΔF_1	0.161	0.178	0.905	.183
ɔ F_1	4.373	0.631	6.936	< .001	ɔ ΔF_1	1.943	0.269	7.215	< .001
aʊ F_1	4.817	0.562	8.579	< .001	aʊ ΔF_1	−0.290	0.174	−1.664	.048
ɑɪ F_1	9.924	0.736	13.482	< .001	ɑɪ ΔF_1	−0.714	0.222	−3.219	.001
ɝ F_1	−6.046	0.551	−10.978	< .001	ɝ ΔF_1	−0.698	0.202	−3.462	< .001
eɪ F_1	−1.902	0.657	−2.896	.002	eɪ ΔF_1	−1.800	0.264	−6.833	< .001
ɪ F_1	−9.099	0.718	−12.667	< .001	ɪ ΔF_1	−0.798	0.263	−3.033	.001
i F_1	−17.565	1.771	−9.915	< .001	i ΔF_1	−0.680	0.559	−1.215	.112
oʊ F_1	−1.222	0.485	−2.519	.006	oʊ ΔF_1	−0.896	0.177	−5.060	< .001
ɔɪ F_1	−7.921	1.013	−7.823	< .001	ɔɪ ΔF_1	−2.639	0.581	−4.538	< .001
ʊ F_1	−5.871	0.813	−7.220	< .001	ʊ ΔF_1	−2.733	0.488	−5.596	< .001
u F_1	−14.323	0.950	−15.080	< .001	u ΔF_1	−1.048	0.331	−3.169	.001
ɑ F_2	−14.352	0.798	−17.995	< .001	ɑ ΔF_2	−0.958	0.340	−2.815	.002
æ F_2	1.379	0.572	2.411	.008	æ ΔF_2	0.372	0.297	1.254	.105
ʌ F_2	−9.886	0.627	−15.760	< .001	ʌ ΔF_2	−0.098	0.297	−0.331	.370
ɔ F_2	−20.339	0.980	−20.759	< .001	ɔ ΔF_2	−2.544	0.396	−6.429	< .001
aʊ F_2	−4.301	0.583	−7.379	< .001	aʊ ΔF_2	−2.936	0.328	−8.948	< .001
ɑɪ F_2	−5.071	0.785	−6.462	< .001	ɑɪ ΔF_2	2.842	0.368	7.713	< .001
ɝ F_2	−3.837	0.524	−7.328	< .001	ɝ ΔF_2	1.039	0.318	3.263	.001
eɪ F_2	3.269	0.604	5.415	< .001	eɪ ΔF_2	4.125	0.405	10.196	< .001
ɪ F_2	4.288	0.627	6.841	< .001	ɪ ΔF_2	1.616	0.406	3.978	< .001
i F_2	11.341	1.428	7.940	< .001	i ΔF_2	2.452	0.699	3.506	< .001
oʊ F_2	−7.130	0.580	−12.290	< .001	oʊ ΔF_2	−2.736	0.317	−8.644	< .001
ɔɪ F_2	−23.496	1.659	−14.161	< .001	ɔɪ ΔF_2	2.387	0.587	4.066	< .001
ʊ F_2	−18.695	1.354	−13.812	< .001	ʊ ΔF_2	1.379	0.536	2.572	.005
u F_2	−2.250	0.685	−3.282	.001	u ΔF_2	−3.714	0.543	−6.839	< .001
ɑ F_3	−0.638	0.329	−1.942	.026					
æ F_3	−0.237	0.347	−0.683	.247					
ʌ F_3	−0.048	0.285	−0.167	.433					
ɔ F_3	1.082	0.363	2.980	.001					
aʊ F_3	0.493	0.344	1.433	.076					
ɑɪ F_3	−0.452	0.376	−1.204	.114					
ɝ F_3	−0.366	0.215	−1.704	.044					
eɪ F_3	0.514	0.361	1.424	.077					
ɪ F_3	−0.111	0.288	−0.385	.350					
i F_3	0.596	0.526	1.133	.129					
oʊ F_3	1.081	0.279	3.880	< .001					
ɔɪ F_3	2.840	0.593	4.787	< .001					
ʊ F_3	1.753	0.445	3.937	< .001					
u F_3	−0.575	0.330	−1.744	.04					

Table 55: The outputs of a multinomial logistic regression for speaker GAI's ADS productions